arXiv:2603.19806v1 [cs.LO] 20 Mar 2026

¨余 Büchi 自动机链有多简洁?

¨RUDIGER EHLERS

克劳斯塔尔工业大学，软件与系统工程研究所，Arnold-Sommerfeld-Straße 1，38678 克劳斯塔尔-采勒费尔德，德国

Abstract. 余 B¨üchi 自动机链（COCOA）最近被引入，作为表示任意 ω-正则语言的一种新型规范模型。它们可以在多项式时间内最小化，因此对于通常使用确定性 ω-自动机的应用而言，是一种颇具吸引力的语言表示形式。虽然已知如何从确定性奇偶自动机构造 COCOA，但目前对它们与早于 COCOA 引入的自动机模型之间的关系知之甚少。

本文分析余 B¨üchi 自动机链的简洁性。我们给出三项主要结果，并概述这些结果的影响。首先，我们证明，即使链中的所有自动机都是确定性的，余 B¨üchi 自动机链也可能比确定性奇偶自动机简洁得多，甚至达到指数级差距。接着，我们给出两项主要结果，二者共同对这样一个问题给出了否定的回答：对 COCOA 执行布尔运算（如析取、合取和补运算）时，能否保持这种简洁性？对于二元运算，我们证明，存在一些语言族，对其应用这些运算会导致自动机大小呈指数级增长。这些语言族具有如下性质：若使用确定性奇偶自动机表示它们，对族中元素取析取或合取只需多项式级膨胀。最后，我们证明，对 COCOA 求补时，指数级膨胀同样不可避免，因为这一运算可能需要重新分配用于识别单词的颜色。

1. Introduction

无限词上的自动机是表示反应式系统规约的经典模型。它们通过为规约提供一种中间表示来增强线性时序逻辑（LTL，[Pnu77]）和线性动态逻辑（LDL，[GV13, FZ17]）等时序逻辑；这种表示经过精心组织，可直接用于验证和综合算法。对于有限状态系统的经典模型检测，带有 B¨uchi 接受条件的非确定性自动机已足够；但在反应式综合和概率模型检测等应用中，则会采用更丰富的自动机类型。在此背景下，带奇偶接受条件的确定性自动机尤其值得关注：若规约以这种形式给出，便可根据自动机的状态空间结构，将针对该规约的反应式综合问题归约为求解一个奇偶博弈 [BCJ18]。

* 本文是发表于 GandALF 2025 [Ehl25a]、题为同名论文的扩展版本。研究经费由大众汽车基金会在其 *Momentum* 框架下提供，项目编号为 9C283。

© R.Ehlers CC BY 知识共享

遗憾的是，从时序逻辑到确定性奇偶自动机的转换过程，在最坏情况下无法避免显著膨胀，这使得它们在反应式综合中的应用更为复杂。例如，对于 LTL 规约，已知其最坏情况膨胀为双指数级 [KR10]。然而，即使对于不需要如此庞大自动机的语言，当前用于生成确定性奇偶自动机的转换过程也可能产生不必要的大型自动机，这是因为这些过程仅采用启发式方法来缩减规模，而确定性奇偶自动机的最小化是 NP 难问题 [Sch10, ARE25],

为了解决这一问题，近期有人提出了共-B¨üchi 自动机链（COCOA），作为表示 ω-正则语言的一种新模型 [ES22a]。在 COCOA 中，要表示的语言被拆分为一个递降的共-B¨üchi 语言链，其中每个共-B¨üchi 语言都由一个带有历史确定性 共-B¨üchi 自动机表示，并采用基于转移的接受条件 （HD-tCBW）。在此语境下，基于转移的接受条件是指转移而非状态被标记为接受或拒绝。这类共-B¨üchi 自动机可以在多项式时间内最小化 [ARK22], ，因此链中的每个自动机都可以单独最小化。为了将这些自动机用于一种适用于任意 ω-正则语言且可在多项式时间内规范化最小的模型，研究者定义了将一个 ω-正则语言规范划分为共-B¨üchi 自动机的方法 [ES22a]。COCOA 可以被理解为给每个词赋予一种颜色，就像确定性奇偶自动机所做的那样。若链中的第 i 个自动机接受某个词，而链中此后没有自动机接受该词，那么该词的颜色就是 i （其中 $i \in \mathbb{N}$）。COCOA 定义的核心贡献在于明确了应为每个词赋予哪种颜色；这种明确方式并不依赖于语言的某种自动机表示，而只取决于语言本身，称为相应词的自然颜色 。

COCOA 已在反应式综合中得到初步应用 [EK24a], ，其依据是将确定性奇偶自动机转换为 COCOA 的过程 [ES22a]。鉴于多项式时间最小化也是未来应用中一项颇具吸引力的特性，因此有必要深入考察 COCOA 的性质及其与早期自动机类型的关系，尤其是与确定性奇偶自动机的关系；在某些应用中，COCOA 有望取代后者。

具体而言，我们需要了解它们相较于以往自动机模型的简洁性，因为这有助于判断在何种情形下，COCOA 能够在验证和综合过程中取代目前所用的其他自动机类型。此外，既然 COCOA 已可进行多项式时间最小化，那么未来利用这种表示处理 ω-正则语言时，或许可以采用组合式方法构建规约模型，并在每一步之后对中间自动机进行最小化。要判断这种方法是否可行，我们需要了解对自动机执行常见布尔运算的复杂度，其中最重要的是合取（语言交集）和补运算。

本文研究了 COCOA 的简洁性。我们将其与确定性奇偶自动机进行比较，并分析对 COCOA 所表示的语言执行布尔运算会如何影响其简洁性。除了总结确定性 co-B¨uchi 自动机的一些已有结果如何适用于 COCOA 之外，我们还给出了三项 COCOA 特有的技术结果：

(1) 我们证明，即使 COCOA 中的 co-B¨uchi 语言可以由小型确定性 co-B¨uchi 自动机表示，且整体语言只有一个剩余语言，COCOA 仍可能比确定性奇偶字自动机（DPW）简洁得多，规模甚至可相差指数级。此前已知 COCOA 可能比确定性奇偶自动机简洁得多，

但这是因为 COCOA 中的自动机具有历史确定性；而已知对于某些语言，HD-tCBW 比确定性自动机简洁得多，规模可相差指数级。本文的新结果表明，即使链元素不具备历史确定性，COCOA 仍可能比 DPW 简洁得多。我们的结果还表明，这些简洁性来源并不能叠加。

(2) 我们证明，COCOA 相对于确定性奇偶自动机所具有的指数简洁性，在对 COCOA 执行二元布尔运算（如合取或析取）时可能会丧失。具体而言，即使在确定性奇偶自动机中无需指数级增长的情形下，这类布尔运算仍可能要求状态数呈指数级增长。

(3) 最后，我们证明，对 COCOA 执行补运算（这是一种仅有单个操作数的布尔运算），在最坏情况下也会导致指数级膨胀。其原因在于，未取补的 COCOA 中具有相同自然颜色的词，在取补后的 COCOA 中可能具有不同的自然颜色。要区分这类词，可能需要在补后 COCOA 的单个链元素中区分原 COCOA 的指数多个剩余语言，从而破坏其简洁性。

这三项结果揭示了 COCOA 的基本性质。第一种情形中，为该结果定义的语言族示例表明，即使链中的自动机规模很小，也能够组合出复杂的活性语言。即使这些自动机没有利用历史确定性余 Büchi 自动机所带来的额外简洁性，这一结论仍然成立。第二项技术结果表明，COCOA 所具有的一项性质——其剩余语言的数量相对于其规模呈指数级增长——在执行二元布尔运算后可能会丧失。因此，这说明未来用于对 COCOA 执行布尔运算的过程，其生成的 COCOA 规模必然存在指数级下界。最后，第三项技术结果表明，对于 COCOA 的某些运算，链中较后出现的余 Büchi 自动机的剩余语言可能会影响输出 COCOA 中第一个余 Büchi 自动机。

本文结构如下：在介绍一些预备知识之后，我们将在第3节概述 COCOA 背后的思路。第4节总结已知结果对 COCOA 简洁性的影响，并给出第一个新的技术结果。随后，第5节给出 COCOA 上二元布尔运算所导致膨胀的下界。关于 COCOA 补运算的第三个技术结果见第6节。最后，本文将在第7节讨论所得结果。

本文是同名论文的扩展版本，原论文发表于2025年9月在马耳他瓦莱塔举行的 GandALF 2025 会议。相比会议版本，本文新增了第6节的结果（关于对 COCOA 取补时产生的指数级膨胀），并补充了引理 5.4、定理 5.7 和引理 5.11 的证明。定理 5.7 的证明还包含新加入的观察 5.8 至 5.10。最后，命题 5.13 的证明也补充了一些细节。

2. Preliminaries

语言： 对于作为字母表的有限集合 Σ，令 Σ^* 表示 Σ上的有限词集合，Σ^ω 表示 Σ上的无限词集合。子集 $L \subseteq \Sigma^\omega$ 也称为语言。给定语言 L 和某个有限词 $w \in \Sigma^*$，我们称 $L|_w = \{w' \in \Sigma^\omega \mid ww' \in L\}$ 是

在上的剩余语言。给定一个词 $w = w_0 w_1 \ldots \in \Sigma^\omega$ 和某个语言 L，如果 $L|_{w_0 \ldots w_i} = L|_{w_0 \ldots w_i \tilde{w}}$，则称无限词 $w' = w_0 w_1 \ldots w_i \tilde{w} w_{i+1} w_{i+2} \ldots$ 是 $\tilde{w}$ 在位置 $i \in \mathbb{N}$ 处通过残余语言不变注入得到的。

自动机： 某些语言，尤其是 ω正则语言，可以用奇偶自动机表示。本文只考虑具有基于转移的接受条件的自动机。这些自动机是如下形式的元组：$\mathcal{A} = (Q, \Sigma, \delta, q_0)$，其中 Q 是一个有限的状态集，Σ 是字母表，$q_0 \in Q$ 是自动机的初始状态，$\delta \subseteq Q \times \Sigma \times Q \times \mathbb{N}$ 是其转移关系。

给定一个词 $w = w_0 w_1 \ldots \in \Sigma^\omega$，若 w 诱导出无限运行 $\pi = \pi_0 \pi_1 \ldots \in Q^\omega$ 以及颜色序列 $\rho = \rho_0 \rho_1 \ldots \in \mathbb{N}^\omega$，则称之为如此；具体而言，$\pi_0 = q_0$ 成立，并且对于所有 $i \in \mathbb{N}$，都有 $(\pi_i, w_i, \pi_{i+1}, \rho_i) \in \delta$。本文只考虑输入完备的自动机，也就是说，对于每个状态/字母组合 (q, x)，至少存在一对 (q', c)，使得 $(q, x, q', c) \in \delta$。此外，我们只考虑颜色 c 不依赖于所采用的转移的自动机，因此对于每个 (q, x)，只有一个值 c 满足：对某个 q'，有 $(q, x, q', c) \in \delta$。

若对应的颜色序列 ρ（该序列唯一）中无穷多次出现的最小颜色为偶数，则称一次运行为接受运行。这个颜色也称为该运行的支配颜色。若存在一条接受运行，则称词 $\mathcal{A}$ 被接受。$\mathcal{A}$的语言记作 $\mathcal{L}(\mathcal{A})$，是所有存在接受运行的词所组成的集合。若对于每个 $(q, x) \in Q \times \Sigma$，恰好存在一个组合 $(q', c) \in Q \times \mathbb{N}$，使得 $(q, x, q', c) \in \delta$，则称自动机具有确定性。在这种情况下，我们也将唯一运行的支配颜色称为自动机识别 该词时所用的颜色。若沿 $\mathcal{A}$ 中的转移出现的颜色只有 1 和 2，则称 $\mathcal{A}$为*co-B¨uchi* 自动机。*co-B¨uchi* 语言 是某个 co-B¨uchi 自动机的语言。

我们称自动机 $\mathcal{A}$ 表示若干自动机 $\mathcal{A}_1$ 和 $\mathcal{A}_2$ 的析取，如果$\mathcal{L}(\mathcal{A}) = \mathcal{L}(\mathcal{A}_1) \cup \mathcal{L}(\mathcal{A}_2)$。如果 $\mathcal{L}(\mathcal{A}) = \mathcal{L}(\mathcal{A}_1) \cap \mathcal{L}(\mathcal{A}_2)$，则称它表示 $\mathcal{A}_1$和 $\mathcal{A}_2$的合取。自动机的大小定义为其状态数。

历史确定自动机： 如上所定义，奇偶自动机不一定是确定性的。我们尤其考虑历史确定性co-B¨uchi 自动机（HD-tCBW）。对于这类自动机，存在某个建议函数 $f\colon \Sigma^* \to Q$，使得对于每个词，当且仅当 $w = w_0 w_1 \ldots \in \mathcal{L}(\mathcal{A})$ 时，序列 $q_0 f(w_0) f(w_1 w_2 \ldots) \ldots$是该自动机的有效接受运行。Abu Radi 和 Kupferman [ARK22] 展示了如何对这类自动机进行最小化；在最小化后的自动机中，对于每种状态/字母组合，所有转移的颜色都相同（因此上述假设是成立的）。文献中也将历史确定性 co-B¨uchi 自动机称为游戏友好型co-B¨uchi 自动机。HD-tCBW 和确定性 co-B¨uchi 自动机所能表示的语言集合相同。然而，确定性奇偶自动机（DPW）的表达能力严格更强。我们有时也用图形表示法表示自动机：状态用圆形表示，转移用圆形之间的箭头表示，初始状态则用从一个点引出的箭头标记。在 co-B¨uchi 自动机中，虚线箭头表示拒绝转移（颜色为 1），实线箭头则表示接受转移（颜色为 2）。对于奇偶自动机，边上除了标注字母表符号外，还标注颜色编号。

强连通分量： 给定一个自动机 $\mathcal{A} = (Q, \Sigma, \delta, q_0)$，如果元组 (Q', δ') 满足$Q' \subseteq Q$ 且 $\delta' \subseteq \delta$，则称其为 $\mathcal{A}$ 的一个强连通分量 （SCC）：对于每个 $q, q' \in Q'$，都存在一条 δ' 内的转移序列，可从 q 到达 q'。类似地，δ' 内的每个转移都出现在这样一条序列中。

时序逻辑与 ω-正则表达式： 线性时序逻辑 （LTL， [Pnu77]）是一种用于表达 $\Sigma = 2^{\mathsf{AP}}$ 上（某些）语言的形式体系，适用于一组 AP。已知 LTL 可以转换为确定性奇偶自动机，其大小相对于 LTL 性质的规模呈双指数增长，而且这种膨胀界是紧的（例如，参见 [EKRS17]）。我们还使用 ω-正则表达式 来描述某些语言。这类表达式通过引入一个表示无限重复的符号来扩展经典正则表达式，即 $^\omega$。

3. A short introduction to chains of co-Büchi automata

余 Büchi 自动机链（COCOA，单复数形式相同）为任意 ω-正则语言提供了一种规范表示。给定字母表Σ 上的语言 L。COCOA 表示 L 的起点，是将 Σ^ω分解为语言链 $L_1 \supset L_2 \supset \ldots \supset L_n$。词 w属于该链所表示的语言，以下记作 $\mathcal{L}(L_1、\ldots\ldots、L_n)$，当且仅当使 $w \in L_i$成立的最大下标 i为偶数或 $w \notin L_1$。每个语言 L_i （对某个 $1 \leq i \leq n$）表示这样一组词：它们相对于 L的自然颜色至少为 i。词的自然颜色是该词归属的最小颜色，而“归属”的定义如下：

定义 3.1 ([ES22a], 定义 1)**.** 设 L 是一个语言，且 $i \in \mathbb{N}$。我们称词 w 在 i 的某个颜色中是适居的，若存在一个注入点序列 $J \subset \mathbb{N}$，使得对于所有通过将残余语言不变词注入 w 中 J的词位置而得到的词 w'，以下情况之一成立：

- w' 在严格小于 i的某个颜色中是适居的，或者
- w 和 w' 均属于 L，且 i 为偶数；或者 w 和 w' 均不属于 L，且 i 为奇数。

请注意，在 $i = 0$ 的情况下，只有第二种情况适用。

词的自然颜色这一概念，将奇偶自动机中的颜色概念推广到了不依赖于表示该语言所用自动机具体选择的形式。上述归纳定义从颜色 0 开始，因此每一层的语言都有唯一确定的定义。

根据这个定义，不仅每个 ω-正则语言 L都唯一确定一个语言链 L_1、...、 L_n，而且对于每个 $1 \leq i \leq n$，语言 L_i 都是一种 co-B¨uchi 语言 [ES22a], ，也就是说，它可以编码为一个 co-B¨uchi 自动机。因此，我们可以用一条历史确定性 co-B¨uchi 自动机链 A_1、...、 A_n来表示语言链 L_1、...、 L_n。这些自动机都可以在多项式时间内最小化并规范化 [ARK22]。由于将 L 表示为一条 co-B¨uchi 语言链L_1、...、 L_n同样是规范的，因此我们得到了 L的规范表示。

有关 COCOA 语言表示的详细信息，可参阅介绍 COCOA 的论文 [ES22a] ，以及该论文概念介绍的录像，其中还提供了更多示例 [ES22b]。

3.1. 一个 COCOA 示例。 图 1 展示了一个由三个自动机组成的 COCOA，它们都以字母表 $\Sigma = \{a, b, c\}$ 为字母表，共同表示某种语言 L。这条链的语言包含含有无限多个 a 字母（其自然颜色为 0）的词，以及满足以下三个条件的词：

- 该词最终只由 b 和 c 组成，
- 最终，每个 b都会紧接着一个 c，并且

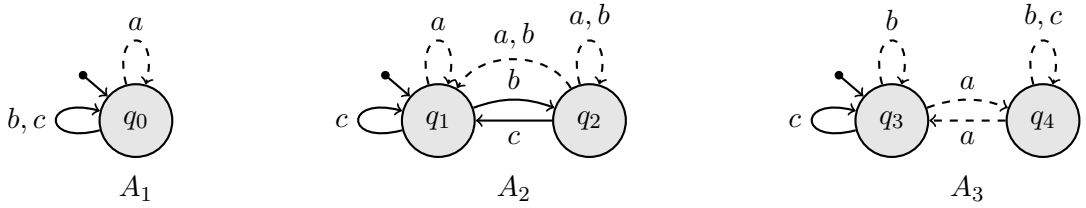

图 1：一个 COCOA 示例

• if th 词中有有限个偶数的 a 字母时，存在无穷多个 b .
满足这三个条件、但只有有限多个 as 的词，其自然颜色为 2。不属于 L 的词，其自然颜色为 1 或 3。因此，COCOA 能识别具有四种不同自然颜色的词；根据从确定性奇偶自动机到 COCOA 的现有转换过程 [ES22a] ，该语言的任何确定性奇偶自动机也至少需要四种颜色。这些颜色表示：通过无限次注入残余语言不变的有限词，词在属于或不属于 L 之间交替的频率。在此例中，词 c^ω 的颜色为 3，因此不属于 L。通过注入 bs，使得所得词中不会出现连续两个 b 字母，该词便属于 L。随后，通过无限次注入 bb，该词又不属于 L 。最后，再通过无限次注入 a ，最终得到的词被 A_1 拒绝，因此属于 L。COCOA 所表示的整体语言 L 有两个剩余语言，但只有 A_3 会对它们进行跟踪，而 A_2 和 A_1 不会。定义 3.1 中注入位置集 J 的重要性无法通过图 1 中的 COCOA 展示，因为对于以下各节讨论的所有具体 COCOA，该集合都可以自由选择，因此并不重要。

3.2. **COCOA 上的一些补充定义和说明**。 为方便起见，在下文讨论 COCOA $A_1, \dots, A_n$ 时，我们一律假设 $A_0 = \Sigma^\omega$ 和 $A_{n+1} = \emptyset$，因为这样可以避免在某些构造中处理特殊情形，同时不会影响 COCOA 的语言定义。为避免下文叙述繁杂，注入 一词始终指保持剩余语言不变的词注入。当词 w' 是通过保持剩余语言不变的词注入得到的，并注入到某个词 w中时，我们称 w' 延伸 w。我们将 COCOA 中各自动机大小之和定义为 COCOA 的大小。

一条共-B¨uchi 自动机链 $A_1, \dots, A_n$ 表示语言 L的条件，可以等价地表述为：要求 A_1 拒绝自然颜色为 0 的单词（相对于 L），并且对于每个 $1 \leq i \leq n$，被 A_i 接受但被 A_{i+1} 拒绝的单词（若$i < n$）恰好是自然颜色为 i 的单词（相对于 L）。上述定义还意味着，词 w 的自然语言只能通过向 w 中注入字母而缩小（若注入位置集合 J 的选取符合定义 3.1 的要求）。

4. On the conciseness of COCOA

本节将比较确定性奇偶自动机与 COCOA 的规模（针对相同的语言）。我们总结现有关于 COCOA 简洁性的结果所带来的推论，并补充新的见解。

COCOA 相较于 DPW 的简洁性。首先，Ehlers 和 Schewe[ES22a] 提出的转换方法，可将具有 n 个状态和 c 种颜色的确定性奇偶自动机转换为至多包含 c 个历史确定性 co-B¨uchi 自动机的 COCOA，且每个自动机至多有 n 个状态。此外，由于 Ehlers 和 Schewe 的构造会即时最小化颜色数量，因此对于 c（即该语言的任何 DPW 所需的最小颜色数）而言，这一结论同样成立。因此，同一语言的 COCOA 相较于确定性奇偶自动机至多大一个多项式因子，而这个因子的上界由颜色数决定。这个界也是紧的：

命题 4.1 (似乎此前未在其他地方陈述过)。令 无穷出现集合 为一个函数，它将序列映射到在该序列中无穷多次出现的元素集合。对于每个 $k \in \mathbb{N}$，语言 $L^k = \{w \in \{1,\dots,k\}^\omega \mid \min(\inf(w)) \text{ 为偶数}\}$，可由一个仅有一个状态、使用 k 种颜色的确定性奇偶自动机表示；但表示同一语言的每个 *COCOA* 至少需要 k 个层级（每层一个状态）。

证明。可以构造一个仅有一个状态的确定性奇偶自动机，为所有字母设置自环，并将字母本身用作相应的颜色。将确定性奇偶字自动机转换为 COCOA 的现有过程 [ES22a] 随后会为该语言构造一个 COCOA A_1，.....，A_k，其中在每个层级 i，以 $(\{i,\dots,k\})^\omega$ 结尾的单词都会被接受。总体而言，膨胀因子为 k，而 k 是初始确定性奇偶自动机中的颜色数。

□

线性时序逻辑 $\to$ COCOA。在讨论 COCOA 也可以比确定性奇偶字自动机更简洁之前，我们先来看一个二者简洁性相同的领域。具体而言，将线性时序逻辑转换为自动机时，确定性奇偶字自动机和 COCOA 的最坏情况膨胀下界相同，即为双指数级。

这可以从一项已有证明中得出：该证明给出了将线性时序逻辑转换为确定性 B¨üchi 自动机（在可行时）的双指数级下界。Kupferman 和 Rosenberg 针对字母表固定和不固定的情形，给出了这类证明的多个版本 [KR10]。所有证明都有一个共同点：它们构造了一个语言族，其剩余语言的数量相对于线性时序逻辑公式的大小呈双指数级增长。这一点可以从以下事实看出：这些语言只包含以某个字母表中的字符 # 结尾的词，因此，只有剩余语言数量发生双指数级膨胀，才能使自动机变得如此庞大。

这些语言的补集可以由 co-B¨uchi 自动机表示。此外，最小的 HD-tCBW 是语义确定的，这意味着对于自动机 $A = (Q, \Sigma, \delta, q_0)$ 中由前缀词 $\tilde{w}$ 可达的每个状态 q，都有 $\mathcal{L}((Q, \Sigma, \delta, q)) = \{w \in \Sigma^\omega \mid \tilde{w}w \in \mathcal{L}(A)\}$。如果 A 中有双指数数量的剩余语言，那么 A 至少需要双指数数量的状态。因此，这些语言对应的 COCOA 也需要双指数规模，因为 co-B¨uchi 语言的 COCOA 仅由该语言的一个 HD-tCBW 构成。

COCOA 相对于确定性奇偶自动机的简洁性。现在我们来探究 COCOA 是否以及如何比确定性奇偶自动机更简洁。首先，已有研究表明，HD-tCBW 的简洁度可以呈指数级优于确定性 co-B¨uchi 自动机 [KS15]。由于奇偶自动机属于 co-B¨uchi 类型 [KMM06], ，确定性 co-B¨uchi 字自动机不可能比确定性奇偶自动机更简洁。此外，由于针对 co-B¨uchi 语言的 COCOA 由

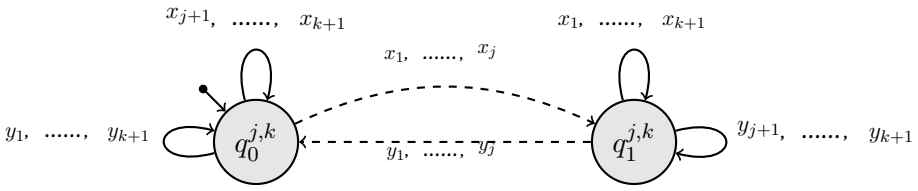

图2：一个确定性 co-B¨uchi 自动机（针对某些 $k \in \mathbb{N}$ 和 $1 \leq j \leq n$ 参数化），用于证明定理 4.3 中所用的 co-B¨uchi 语言

单个历史确定性 co-B¨uchi 自动机构成，因此总的来说，COCOA 的简洁度可以呈指数级优于确定性奇偶自动机。

我们还可以利用一些已有结果来证明，COCOA 不可能比确定性奇偶自动机简洁超过双指数级：

命题 4.2 (此前已在 [EK24b] 中以简略形式出现，其依据是[EK24a]中的评述)。 设 $(A_1,\ \ldots\ldots,\ A_k)$ 是一个 *COCOA*。存在一个识别相同语言的确定性奇偶自动机，其状态数相对于 $|A_1| + \ldots + |A_k|$ 呈指数级增长。

证明。 将非确定性 co-B¨uchi 自动机转换为确定性 co-B¨uchi 自动机，可以通过 Miyano-Hayashi 构造 [MH84] 实现，所需代价为指数级膨胀 [BKR10] 。对 COCOA 中的每个自动机执行此操作，可得到一个确定性 co-B¨uchi 自动机序列 $\mathcal{D}_1,\ \ldots\ldots,\ \mathcal{D}_k$，其中对每个 $1 \leq j \leq k$，我们都有$|\mathcal{D}_j| \leq 3^{|A_j|}$。

对每个 $1 \leq j \leq k$，令其为 $\mathcal{D}_j = (Q^j,\ \Sigma,\ \delta^j,\ q_0^j)$。我们可以如下构造一个确定性奇偶自动机 $\mathcal{P} = (Q^P,\ \Sigma,\ \delta^P,\ q_0^P)$，使其识别 COCOA 的语言（使用 [EK24b] 中的构造）：

$$
\begin{aligned}
Q^P &= Q^1 \times \ldots \times Q^k \\
\delta^P((q^1, \ldots, q^k), x) &= ((q'^1, \ldots, q'^k), c) \text{ s.t. } \exists c^1, \ldots, c^k \in \mathbb{N}.(q'^1, c^1) \in \delta^1(q^1, x), \ldots, \\
&\qquad (q'^k, c^k) \in \delta^k(q^k, x), c = \min(\{k\} \cup \{j \in \{0, \ldots, k-1\} \mid c^{j+1} = 2\}) \\
q_0^P &= (q_0^1, \ldots, q_0^k)
\end{aligned}
$$

为了证明 $\mathcal{P}$ 具有正确的语言，假设对于某个词 w，其自然颜色是 j，其中 $0 \leq j \leq k$。那么，所有自动机 $\mathcal{D}_1,\ \ldots\ldots,\ \mathcal{D}_j$都接受该词，而自动机$\mathcal{D}_{j+1},\ \ldots\ldots,\ \mathcal{D}_k$都拒绝该词。由于 $\mathcal{P}$并行模拟了所有这些自动机，因此在 w 的运行中，沿转移经过颜色 c 的情况会无穷多次出现，而颜色属于 $\{0, \ldots, j-1\}$的情况只会出现有限多次。这意味着 $\mathcal{P}$接受 w当且仅当 j是偶数，由此证明 $\mathcal{P}$ 具有正确的语言。

由此可得 $|\mathcal{P}| \leq |\mathcal{D}_1| \cdot \ldots \cdot |\mathcal{D}_k| \leq 3^{|A_1|} \cdot \ldots \cdot 3^{|A_k|} = 3^{|A_1| + \ldots + |A_k|}$。总的来说，翻译的膨胀因此是指数级的。 □

乍看之下，COCOA 相较于确定性奇偶字自动机（DPW）的简洁性已被精确地刻画为单指数级。然而，现有结果无法轻易推导出的是，COCOA 究竟为何会比确定性奇偶自动机简洁一个指数级。尤其是，除历史确定性 co-B¨uchi 自动机的简洁性之外，可能还有其他因素促成了 COCOA 的简洁性，但这些因素并未叠加。

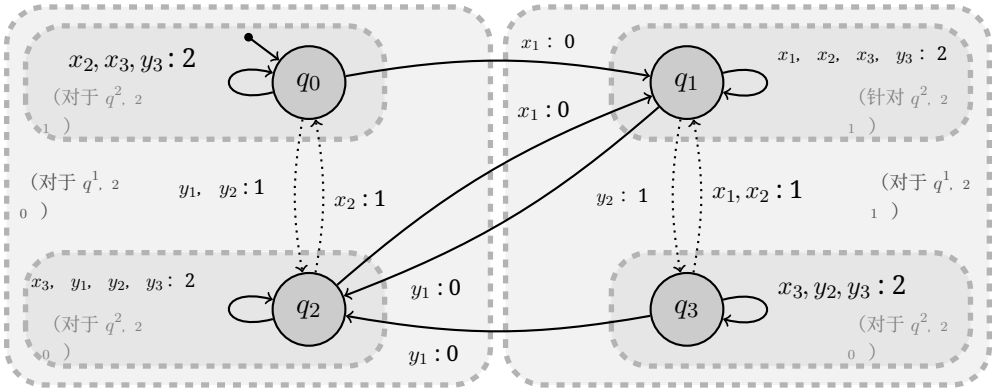

图 3：一个用于 $\mathcal{L}(\mathcal{C}^2)$的最小DPW，并标示出各状态如何映射到同一语言的 COCOA 中的状态组合。颜色为 1 的转移以虚线绘制。

事实确实如此，下面我们将对此加以说明。即使 COCOA 每一层上的 co-B¨uchi 语言都可以表示为两状态的确定性 co-B¨uchi 自动机，表示该语言的奇偶自动机仍可能需要多出指数级数量的状态。

定理 4.3。 存在一个 *COCOA* 族 $\mathcal{C}^1, \mathcal{C}^2, \ldots$，使得对于其中每个 *COCOA* $\mathcal{C}^k = (A_1^k, \ldots, A_k^k)$，对所有 $1 \leq j \leq k$，历史确定性 *co-B¨uchi* 自动机 A_j^k 仅有两个状态，$\mathcal{C}^k$ 的语言仅有一个剩余语言，而该语言的每个确定性奇偶自动机 $\mathcal{P}^k$ 至少需要 2^k 个状态。

为了证明这个定理，我们首先定义一个合适的语言族。

定义4.4。 对于每个 $k \in \mathbb{N}$，我们定义 $\mathcal{C}^k = (A_1^k, \ldots, A_k^k)$，使 COCOA 中的 co-B¨uchi 自动机具有联合字母表 $\Sigma^k = \{x_1, \ldots, x_{k+1}, y_1, \ldots, y_{k+1}\}$，并且对每个 $1 \leq j \leq k$，都可以如图2所示给出一个识别 A_j^k 的确定性 co-B¨uchi 自动机。

直观而言，COCOA $\mathcal{C}^k$ 中的每个自动机 A_i^k 都会接受这样的词：其中字母 $x_1 \ldots x_i$ 只出现有限次，或者字母 $y_1 \ldots y_i$ 只出现有限次。图 3 展示了一个针对 $\mathcal{L}(\mathcal{C}^2)$ 的最小规模确定性奇偶字自动机 $\mathcal{P}^2$，并有助于理解为什么这类 COCOA 的确定性奇偶字自动机可能需要很大：为了确保确定性奇偶字自动机接受那些被 A_1^2拒绝的词，其状态集必须划分为对应于状态 $q^{1,\ 2\,0}$（左侧）的状态，以及对应于 $q^{1,\ 2\,1}$（右侧）的状态，从而在这两部分之间无限次切换的运行能够被接受。这是通过在确定性奇偶字自动机左右两部分之间的转移上赋予颜色 0 来实现的，而 0 随后成为该运行的支配颜色。然而，在每个 这样的独立状态集中，我们还需要将对应于 $q^{2,\ 2\,0}$（确定性奇偶字自动机中底部的两个状态）的状态，与对应于 $q^{2,\ 2\,1}$（确定性奇偶字自动机中顶部的两个状态）的状态区分开，以检测某个词何时应被确定性奇偶字自动机拒绝，因为它未被 A_2^2接受。底部状态与顶部状态之间的转移颜色为 1，以此确保：如果该词被 A_2^2 拒绝（但被 A_1^2接受），那么该词也会被 $\mathcal{P}^2$ 拒绝。这种将不同 co-B¨uchi 自动机中的状态嵌套起来的方式在 $\mathcal{C}^k$ 中确实不可避免；我们接下来将证明这一点，以证明定理 4.3。

我们借鉴了对重新铺轨自动机 [Ehl25b], 的研究思路，这类自动机是确定性奇偶自动机的推广。具体而言，我们研究奇偶自动机中的强连通分量

在 $\mathcal{L}(\mathcal{C}^k)$ 中是否需要嵌套。通过这种方式可以得到、并用于证明定理 4.3 的主要观察结果，体现在以下引理中：

引理 4.5。设 (Q', δ') 是 $\mathcal{P}^k$ 中的一个强连通分量，仅由可达状态组成，并且对于某些 $1 \leq i < k$ 和 $1 \leq j < k$，我们有：对于任意词 w，其中只出现来自 x_i、……、x_{k+1} 以及 y_j、……、y_{k+1} 的字母，从 Q' 中任意状态开始的关于 w 的运行始终停留在 (Q', δ') 中。

于是，我们有：在(Q', δ') 内存在不相交的可达强连通分量 (Q'^x, δ'^x) 和 (Q'^y, δ'^y)，使得

- 对于任意词 w'，其中只出现来自 $x_{\max(i,j)+1} \dots x_{k+1}$ 以及 y_j、……、y_{k+1} 的字母，从状态 $q \in Q'^x$ 出发的关于 w' 的任意运行都停留在 (Q'^x, δ'^x) 中，并且
- 对于仅包含来自 $x_i \dots x_{k+1}$ 的字母的任意词 w'，以及 $y_{\max(i,j)+1}, \dots, y_{k+1}$，从状态 $q \in Q'^y$ 出发、针对 w' 的任意运行都会停留在 (Q'^y, δ'^y) 中。

证明。我们可以如下找到强连通分量 (Q'^x, δ'^x)：考虑 (Q', δ') 中针对来自 $x_{\max(i,j)+1} \dots x_{k+1}, y_j, \dots, y_{k+1}$ 的字母的转移集 T^x。我们选取 T^x 的一个子集 δ'^x，使其构成某个强连通分量的转移集。这样的子集必定存在，因为 Q' 中状态针对所考虑字母集中的字母所进行的所有转移都留在 Q' 中，而 Q' 与 T^x 一起分解为若干强连通分量。我们以相同的方式找到强连通分量 (Q'^y, δ'^y)，但这次针对字母 $x_i \dots x_{k+1}, y_{\max(i,j)+1}, \dots, y_{k+1}$。

强连通分量 (Q'^x, δ'^x) 和 (Q'^y, δ'^y) 具有所需性质：对于所考虑字符集中的字母，所有出转移分别都在 δ'^x/δ'^y 内，因为它们由强连通分量中相应字符的所有转移组成；此外，由于这些强连通分量的选取方式，在 $\mathcal{P}^k$ 中，相应字母集没有出转移。

要证明 (Q'^x, δ'^x) 和 (Q'^y, δ'^y) 不相交，先考虑一个词 w^x，它包含来自 $x_{\max(i,j)+1} \dots x_{k+1}, y_j, \dots, y_{k+1}$ 的所有字母，且这些字母都无穷次出现；并且从某个 $q'^x \in Q'^x$ 开始，w^x 的一次运行会无穷次地经过 δ'^x 中的所有转移。由于 (Q'^x, δ'^x) 是一个强连通分量，且包含这些字母对应的转移，因此这样的词必定存在。注意，根据 $\mathcal{C}^k$ 的定义，w^x 属于 $\mathcal{C}^k$ 的语言，当且仅当 $\max(i, j)$ 是偶数。我们可以为 $x_i \dots x_{k+1}$ 构造一个类似的词 $w^y, y_{\max(i,j)+1}, \dots, y_{k+1}$。当且仅当 $\max(i, j)$ 是偶数时，它也属于 $\mathcal{C}^k$ 的语言。

如果（Q'^x，δ'^x）和（Q'^y，δ'^y）发生重叠，我们就可以通过以下方式构造词/运行组合w^{mix}/π^{mix}：从 w^x 和 w^y 出发，取 w^x 的前缀运行/词，直到到达共同状态$q_{mix} \in Q'^x \cap Q'^y$；移除 w^y 的词干（即相应运行到达 q_{mix}之前的那些字符）；然后沿着 w^{mix}的运行，每当到达 q_{mix} 时就在两个词之间切换。由此得到的词 w^{mix} 包含 $x_i \dots x_{k+1}$，y_j、……、y_{k+1} 中的所有字母，且每个字母都出现无穷多次；该词对应的运行也会无穷多次经过 $\delta'^x \cup \delta'^y$ 中的所有转移。这意味着，w^{mix} 的运行的支配颜色，等于分别由 w^x 和 w^y诱导的运行的最小主导颜色。

根据 $\mathcal{C}^k$的定义，w^{mix} 是否属于 $\mathcal{C}^k$ 的语言，与 w^x 和 w^y 是否属于 $\mathcal{C}^k$的语言，必然不同，因为当且仅当$\max(i, j)$为奇数时，w^{mix} 才属于 $\mathcal{C}^k$。因此，为避免 w^{mix}、w^x或 w^y 被赋予奇偶性错误的颜色来识别，我们必须使 Q'^x 和 Q'^y 不相交。 □

这个引理可用于针对 $\mathcal{P}^k$ 大小进行的归纳论证：

引理 4.6. 设(Q', δ') 是 $\mathcal{P}$ 中的一个强连通分量，满足：对于任意状态$q \in Q'$，以及仅由 $x_i, \dots, x_{k+1}, y_j, \dots, y_{k+1}$ 中的字母组成的任意词 w，从 q 开始的 w 的运行始终留在 (Q', δ') 中（对某些 $1 \leq i \leq k$ 和 $1 \leq j \leq k$）。我们有，Q' 的大小至少为 $2^{k-\max(i,j)}$。

证明. 我们通过对 $\max(i,j)$ 使用归纳法来证明该断言，从 $\max(i,j) = k$ 的情形开始，并逐步向后推。对于归纳基础（$\max(i,j) = k$），该断言显然成立，因为 (Q', δ') 中至少需要一个状态。

对于归纳步骤，考虑（i，j）与 $i < k$ 和 $j < k$ 的一个具体组合（因此归纳基础不适用）。引理 4.5 指出，在（Q'，δ'）中，分别存在由字母 $x_{\max(i,j)+1} \cdots x_{k+1}$，$y_j$，...，$y_{k+1}$ 和 $x_i \ldots x_{k+1}$，$y_{\max(i,j)+1}$，...，y_{k+1} 对应的不同强连通分量子集（Q'^x，δ'^x）和（Q'^y，δ'^y）。根据归纳假设，它们的大小均为 $2^{k-\max(i,j)-1}$。由于（Q'，δ'）将这两个强连通分量子集都作为不同的子集，因此 Q' 至少需要 $2^{k-\max(i,j)}$ 个状态。 □

我们现在可以证明定理 4.3 了。请注意，我们尚未证明 $\mathcal{C}^k$确实是其所表示语言的规范 COCOA；这要求每个词相对于 $\mathcal{C}^k$ 的语言都以其自然颜色被接受。因此，下面的证明将从确立这一事实开始。

定理 *4.3* 的证明。我们先证明，对于每个$k \in \mathbb{N}$，$\mathcal{C}^k$ 都是某个语言的 COCOA。为此，先考虑一个被 A_1^k拒绝的词 w。那么，x_1 和y_1 都会在该词中无限次出现。注入额外的字母不会改变该词被拒绝这一事实，因此被 A_1^k 拒绝的词具有自然颜色 0。

对于其他自动机，我们通过归纳法证明：若自动机 A_i^k 是接受某个词 w的最小索引自动机，则该词相对于 $\mathcal{C}^k$的语言具有自然颜色 i。因此，假设 w 被 A_i^k 接受，但被 A_{i+1}^k 拒绝（若 $i < k$）。那么，该词要么包含 x_{i+1}无限多次，且所有字符 x_1、.....、x_i都只出现有限次；要么包含 y_{i+1}无限多次，且所有字符 y_1、.....、y_i都只出现有限次。任意注入要么保持这一性质（从而保持该词是否属于 $\mathcal{C}^k$的语言），要么无限多次注入来自 x_1、.....、x_i或 y_1、.....、y_i的字符，此时得到的词具有严格更小的自然颜色。

为了证明规模界，对 x_1、.....、x_{k+1}、y_1、.....、y_{k+1} 以及 (Q, δ) 中任意一个没有出边的强连通分量应用引理 4.6，可得至少需要 2^k 个状态来构造 $\mathcal{P}^k$。注意，这样的强连通分量总是存在。 □

我们注意到，对于本节定义的语言族，定理 4.3 的规模界实际上是紧的，因为通过推广图 3 所示的构造，我们可以得到大小恰为 2^k 的奇偶自动机 $\mathcal{P}^k$。

5. COCOA disjunction/conjunction can cause an exponential blow-up

我们在上一节中看到，与确定性奇偶自动机相比，COCOA 可以做到指数级精简。但这种精简有多脆弱 呢？具体来说，是否可能出现这样的情况：某个语言可以用 COCOA 精简地表示（与 DPW 表示相比），但在处理该语言时，对语言执行的操作却会破坏这种精简？事实证明，对 COCOA 执行合取和析取操作时确实会出现这种情况，正如我们将在本节中展示的那样。我们定义了两个可精简表示的语言族，并证明对该语言族取合取或析取时，不可避免地会产生指数级膨胀。相比之下，使用 DPW 表示该语言族时，合取或析取操作只会导致多项式级膨胀。

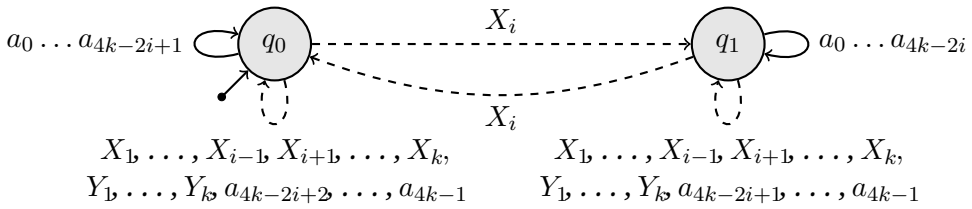

图 4：语言 L_i^k 的确定性 co-B¨uchi 自动机。拒绝转移以虚线表示。

我们指出，这种膨胀与对构成 COCOA 的 HD-tCBW 执行析取或合取操作可能导致的自动机大小增长无关。相反，精简程度的变化是由于重新调整了待表示语言映射到 COCOA 层级的方式。我们还指出，在一般情况下，对 DPW 取合取或析取会不可避免地产生指数级膨胀 [Bok18]。

我们首先介绍第一个语言族 $\{L^k\}_{k\in\mathbb{N}}$，其 COCOA 大小关于 k 为多项式，但其剩余语言的数量关于 k 为指数级。

定义 5.1. 设 $k \in \mathbb{N}$ 给定。我们令 $\Sigma = \{X_1, \dots, X_k, Y_1, \dots, Y_k, a_0, \dots, a_{4k-1}\}$，并针对以下语言序列 $L_1^k, \dots, L_k^k$ 定义 $L^k = \mathcal{L}(L_1^k, \dots, L_n^k)$，其中 $1 \le i \le k$:

$$L_i^k = ((\Sigma \setminus \{X_i\}) + X_i(\Sigma \setminus \{X_i\})^* X_i)^*(a_0 + \ldots + a_{4k-2i+1})^\omega + \Sigma^*(a_0 + \ldots + a_{4k-2i})^\omega$$

每个语言 L_i^k 只包含最终仅由小写字母组成的单词。哪些这样的单词属于该语言，取决于哪些小写字母出现无穷多次，而它们的顺序并不重要。如果单词开头的 X_i 个字母数量为偶数，那么该单词中可以无穷多次出现的字符集合会稍大一些，因为还包括 $a_{4k-2i+1}$。对于所有 L_i^k，可能无穷多次出现的字母集合都严格大于 L_{i+1}^k 的情况。单词中 X_i 个字母的数量是偶数还是奇数，只与 L_i^k 有关，而与 L_j^k 对于 $i \neq j$ 无关。接下来我们将证明

•每种语言 L_i^k 都是 co-B¨uchi 语言，并且存在一个确定性 co-B¨uchi 自动机，识别 L_i^k，且有 2 个状态，并且
• Σ^ω 包含具有 $0 \dots k$ 自然颜色的单词，并且 L_i^k恰好接受具有 i 或更多自然颜色的单词（相对于 L^k）——因此，co-B¨uchi 自动机 L_1^k、.....、 L_k^k 共同构成一个有效的 COCOA。

引理 5.2. 设 L_i^k 是定义 *5.1* 中定义的语言。存在一个具有基于转移的接受条件的确定性 *co-B¨uchi* 自动机，识别 L_i^k，且有两个状态。

证明。 图 4 所示的自动机接受所需的语言。 □

我们注意到，对于同一语言，不存在更小的历史确定性（或确定性）自动机，因为该自动机中有两个剩余语言：状态 q_0 的语言包含 $(a_{4k-2i+1})^\omega$，而状态 q_1 的语言则不包含。Abu Radi 和 Kupferman 已证明，每个 co-B¨uchi 语言都有一个最小的 HD-tCBW，其中每个状态都以其剩余语言 [ARK22]，标记，因此我们由此得知，该语言不存在只有一个状态的 HD-tCBW。

这个定义的一个有趣性质是，由语言链诱导的整体 ω-正则语言，其剩余语言的数量相对于链中的自动机状态数呈指数增长。

引理 5.3。 L^k 的剩余语言数量为 2^k。特别地，对于每一种情况——字母 X_i 迄今出现的次数为偶数还是奇数（对于 $1 \leq i \leq k$）——都恰有一个剩余语言。

证明。为证明该断言，我们需要说明，对于任意两个前缀词 $w'^1 \in \Sigma^*$ 和$w'^2 \in \Sigma^*$，如果 w'^1 和 w'^2 在 X_i 字符（对于某个$1 \leq i \leq k$）出现次数的奇偶性上不同，那么存在一个无限词 $w \in \Sigma^\omega$，使得 $w'^1 w \in L^k$ 是否成立与 $w'^2 w \in L^k$是否成立不同。

取 $1 \leq i \leq n$，不失一般性地设w'^1 中 X_i 字符的数量为偶数，而 w'^2 中 X_i 字符的数量为奇数。那么， $w =(a_{4k-2i+1})^\omega$ 就是这样一个词。根据 L_i^k和 L_{i+1}^k的定义， $w'^1 w \in L^k$ 当且仅当 i 为偶数。根据 L_i^k和 L_{i-1}^k的定义， $w'^2 w \in L^k$ 当且仅当 i 为奇数。 □

在定义用于与 L^k 结合的 COCOA 之前，我们需要证明，对于 $L_1^k, \ldots, L_k^k$，一个共 B¨uchi 自动机序列确实是 L^k的 COCOA，即它能够识别每个具有其自然颜色的词。

引理 5.4。 设 $k \in \mathbb{N}$和 L_1^k，…，L_k^k 是根据定义 *5.1* 为 $k \in \mathbb{N}$ 定义的语言。则每个语言 L_i^k 恰好包含相对于 L^k具有 i自然颜色的词。

证明。首先注意，对于每个 $i \in \mathbb{N}$，都有 $L_i^k \supset L_{i+1}^k$（如果 $i < k$）。这是因为 L_i^k包含所有以（$a_0 + \ldots + a_{4k-2i}$）$^\omega$结尾的词，而 L_{i+1}^k中包含的所有词都必须以（$a_0 + \ldots + a_{4k-2(i-1)+1}$）$^\omega$结尾。因此，我们有 $L_1^k \supset \ldots \supset L_k^k$。

现在我们通过归纳法证明：对于每个 $1 \leq i \leq k$，语言 L_i^k恰好包含关于 L^k 的自然颜色至少为 i的词。

归纳基础： *</style}*令 为一个词。我们需要证明，恰好那些词 满足：对于每个导致某个词 的注入序列，若该词也具有 ，则这些词不在 L_1^k中。由于所有相关语言只关注一个词中出现的字母，而不关注这些字母在词中的位置，因此我们无需针对残余语言不变词被注入 w的位置索引集合进行推理。

⇒: 因此，令 w 为一个不在 L_1^k中的词。我们需要证明，对于一个w的扩展词 w'，同样有 $w' \in L^k$。我们根据可能导致 $w \notin L_1^k$ 的原因进行分类。

• 如果 w中有来自 X_1、.....、X_k、Y_1、.....、Y_k的无穷多个字母，那么注入额外的字母不会改变这一点，因此 w' 也不在 L_1^k 中，这意味着$w' \in L^k$。

• 如果 w 中有无穷多个字母 a_{4k-1}，且前一种情况不成立，那么 w 中就有奇数个 A_1 字母。在 w' 中，要么有无穷多个来自$X_1, \ldots, X_k, Y_1, \ldots, Y_k$的字母被注入，此时根据上述推理， w' 属于 L^k；要么只有有限多个这样的字母被注入，但有偶数个 A_1 字母（否则某个注入就不是残余语言不变的）。在这种情况下，由于 a_{4k-1} 仍然在 w' 中无穷多次出现，且 w' 中有奇数个 X_i 字母，我们也有 $w' \notin L^k$。由于所有其他词都在 L^k

[1], 归纳基础的这一方向到此完成。

⇐: 设 w 是 L^k 中的一个词，且对每个扩展 w 的词 w'，我们也有 $w' \in L^k$。我们必须证明此时 $w \notin L_1^k$ 成立。

设 w 是这样一个词。如果 w 中包含无穷多个 X_i 或 Y_i 字母（对于某个 $1 \leq i \leq k$），那么 w 和 w' 都不属于 L_1^k。在所有其他情形下，我们可以假设该词中 X_1 字母的数量为奇数，且包含无穷多个 a_{4k-1} 字母；否则，通过将无穷多个 a_{4k-2} 字母注入 w，我们可以确保 $w' \in L_1^k$ 成立而 $w' \in L_2^k$ 不成立，这与前提矛盾。这样的词 w 也会被 L_1^k 拒绝，从而完成归纳基础的这一部分。

归纳步骤：在归纳步骤中，我们需要证明：对于 $i \geq 2$，假设 L_{i-1} 恰好接受相对于 L^k自然颜色至少为 $i-1$ 的单词，那么 L_i^k 不包含自然颜色小于 i 的单词 w。这等价于说， L_i^k 恰好拒绝满足以下条件的单词 w：对于每个扩展，要么 (a) 得到的单词 w' 属于 L^k 当且仅当 w 属于其中；要么 (b) 得到的单词的自然颜色小于 i。

$\Rightarrow$：令 $w \notin L_i^k$。我们需要证明，对于 w 的每个扩展，要么 (a) 得到的单词 w' 属于 L^k 当且仅当 w 属于其中；要么 (b) 得到的单词的自然颜色小于i。

设这样的单词 w 已给定，并考虑一个单词 w'。我们进行分类讨论 t:

• 如果 w' 包含无穷多个形如 X_i 或 Y_i 的字母（对于某个 $1 \leq i \leq k$），那么它不会被 L_i^k 接受，也不会被 L_1^k接受。根据归纳假设，断言在此情况下成立，因为该词的自然颜色为 0。

• 如果 w' 不包含无穷多个形如 X_i 或 Y_i 的字母（对于某个 $1 \leq i \leq k$），那么 w也不包含。令 a_h 为在 w 中无穷多次出现且具有最高指标 h 的字母，并令 $a_{h'}$ 为在 w'中无穷多次出现且具有最高指标 h' 的字母。每个大写字母 X_i 只能被注入偶数次，否则 w' 将无法扩展 w （见引理 5.3）。在此情况下，由于为某个 $0 \leq l \leq 4k-1$注入额外的 a_l 字母不会使一个不属于 L_i^k 的词变为包含于 L_i^k中的词，根据归纳假设，w' 要么具有低于 i的自然颜色，要么有 $w' \in L_{i-1}^k$，进而有 $w \in L^k$ 当且仅当 $w' \in L^k$。

$\Leftarrow$：设 w 是一个词，对于由剩余语言不变词注入得到的每个词 w'，要么 w' 属于 L^k 当且仅当 $w \in L^k$，要么（b）所得的词的自然颜色严格小于 i。我们需要证明此时 $w \notin L_i^k$。

设 w 是这样一个词。如果 w 包含无限多个 X_i 或 Y_i 字母（对于某个 $1 \leq i \leq k$），那么它属于 L^k，w'也属于其中。对于所有其他词，我们按最高指标 h分类讨论，其中 w 包含无限多个字母 a_h。

• 如果 $h \geq 4k-2i+2$，那么 L_i^k既不包含 w，也不包含 w'。
• 如果 $h = 4k-2i+1$，且 w 中的 X_i 字母数量为奇数，那么 L_i^k也不包含 w或 w'。

• 假设 $h = 4k-2i+1$，且 w 中的 X_i 字母数量为偶数，那么我们可以通过添加无限多个 a_{h+1} 字母，将 w 扩展为 w'。这使得 $w \in L_i^k$和 $w' \notin L_i^k$。由于$\{w, w'\} \subseteq L_{i-1}^k$，我们有 $(w \in L^k) \neq (w' \in L^k)$。然而，由于 $w' \in L_{i-1}^k$，根据归纳假设， w' 的自然颜色不可能严格小于 i。因此， $h = 4k-2+1$ 且 w 中的 X_i 字母数量为偶数这一假设必定不成立。

• 假设 $h \leq 4k-2i$。那么，如果 h 是奇数，我们可以将无限多个字母 a_{h+1} 注入 w；如果 h 是偶数，则可以将无限多个字母 a_{h+2} 注入其中。根据语言 $L_1^k, \ldots, L_k^k$ 和 L^k 的定义，这些注入会改变该词是否属于 L^k。然而， w 及其扩展都属于

L^k_{i-1} 在这种情况下，这与前提矛盾：每个与 w 在是否包含于 L^k 这一点上不同的 w 的扩展，其自然颜色都小于 i。 □

现在，我们来定义一个语言族，以便与 COCOA 结合使用 L^k 。

定义 5.5。 给定 $k \in \mathbb{N}$。我们令 $\Sigma = \{X_1, \ldots, X_k, Y_1, \ldots, Y_k, a_0, \ldots, a_{4k-1}\}$，并针对以下语言序列定义 $\hat{}L^k = \mathcal{L}(\hat{}L^k_1, \ldots\ldots, \hat{}L^k_n)$，其中 $1 \le i \le k$:

$$\hat{}L^k_i = \Sigma \setminus \{Y_i\}) + Y_i(\Sigma \setminus \{Y_i\})^* Y_i)^*(a_{2i-2} + \ldots + a_{4k-1})^\omega + \Sigma^*(a_{2i-1} + \ldots + a_{4k-1})^\omega$$

((

请注意，引理 5.2 和引理 5.4 中确立的 L^k 的性质对 $\hat{}L^k$ 同样成立，因为这些语言仅通过交换字母 $\{X_i\}_{1\le 1\le k}$ 和$\{Y_i\}_{1\le i\le k}$ 的作用，以及对每个 $0 \le i < 2k$ 交换字母 a_i 和 a_{4k-i-1} 而有所不同。

设下文中的 $L'^k = L^k \cap \hat{}L^k$。接下来，我们将分析 L'^k 所需 COCOA 的规模，从而揭示 L^k 和 $\hat{}L^k$的 COCOA 合取需要多大规模。为进行这项分析，我们考虑语言交集 $L^k_i \cap \hat{}L^k_j$ （针对$1 \le i \le k$ 和 $1 \le j \le k$），并展示如何从一些针对 $L^k_i \cap \hat{}L^k_j$的 co-B¨uchi 自动机的析取中构造出 L'^k 的 COCOA。

引理 5.6。 设 $w \in \Sigma^\omega$ 是一个词。存在唯一的最大指标对 $(i, j) \in \{0, \ldots, k\}^2$，使得 $w \in L^k_i \cap \hat{}L^k_j$，也就是说，$w \in L^k_i \cap \hat{}L^k_j$，并且对于所有满足 $w \in L^k_{i'} \cap \hat{}L^k_{j'}$的 $(i', j') \in \{0, \ldots, k\}^2$，都有 $i' \le i$ 和 $j' \le j$。

此外，对于每一对(i', j') 满足 $i' \le i$ 且 $j' \le j$的情形，都存在 w 的一个扩展 w'，使得(i', j') 是满足 $w' \in L^k_{i'} \cap \hat{}L^k_{j'}$的唯一最大指标对。

证明。对于第一部分，首先注意，根据定义， $w \in L^k_0 \cap \hat{}L^k_0$，因为 L^k_0 和 $\hat{}L^k_0$都包含 Σ上的所有无限词。接着，令 K 为满足$w \in L^k_i \cap \hat{}L^k_j$的元素 (i, j) 的集合。如果对于某些这样的指标对，我们有 $(i, j) \in K$ 且 $(i', j') \in K$，这意味着$w \in L^k_i$, $w \in \hat{}L^k_j$, $w \in L^k_{i'}$，并且 $w \in \hat{}L^k_{j'}$，因此我们也有 $(\max(i, i'), \max(j, j')) \in K$。所以，不可能在 K中同时存在两个关于逐元素比较不可比的极大元素 (i, j) 和 (i', j')；否则，$(\max(i, i'), \max(j, j'))$ 将是 K中的另一个元素，这与 (i, j) 和 (i', j') 都是 K的不可比极大元素这一假设矛盾。

对于断言的后半部分，设给定 w ，令 (i, j) 为 K 中唯一的最大层级，并令 (i', j') 满足 $i' \le i$ 和 $j' \le j$。将 $a_{4k-2i'}$ 无限多次注入 w，所得的词属于 $L^k_{i'}$ （但不属于 $L^k_{i'+1}$）；将 $a_{2j'-1}$ 无限多次注入，所得的词属于 $\hat{}L^k_{j'}$ （但不属于 $\hat{}L^k_{j'+1}$）。定义 5.1 和 5.5 保证，前一类字母注入不会影响所得词在链 $\hat{}L^k_1, \ldots, \hat{}L^k_k$ 中的位置，而后一类字母注入不会影响所得词在链 $L^k_1, \ldots, L^k_k$ 中的位置。因此，该扩展词的唯一最大指标对为 (i', j')。

□

定理 5.7。 $L'^k = L^k \cap \hat{}L^k$ 的 *COCOA* 可表示为 $\mathcal{C}^k = (C^k_1, \ldots, C^k_{2k})$，其中对于每个 $u \in \{0, \ldots, 2k\}$， C^k_u 的语言为

$$\mathcal{L}(C^k_u) = \bigcup_{(i,j)\in\Gamma_u} L^k_i \cap \hat{L}^k_j$$

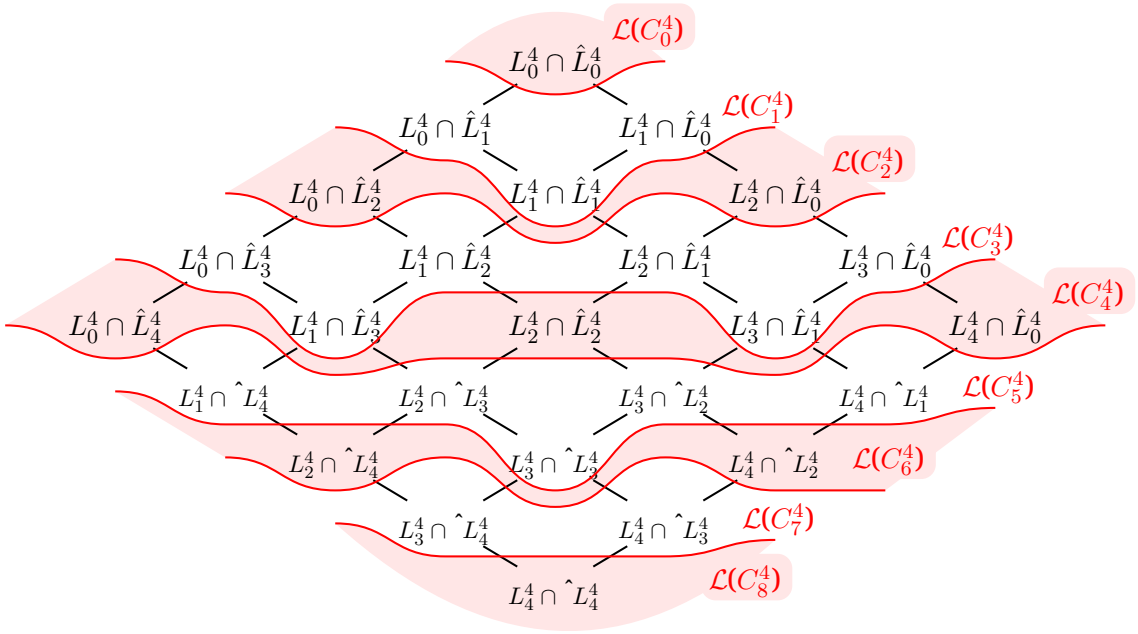

图5：定理 5.7 中集合 $\{L_i^k \cap \hat{L}_j^k\}_{0\le i\le k,\ 0\le j\le k}$（对于 $k=4$）如何组合为 C_0^k、……、C_{2k}^k 的概览

for

$$\Gamma_u^k = \begin{cases} \{(i,j)\in\{0,\dots,k\}^2 \mid i+j=u, i \text{ is even}, j \text{ is even}\} & \text{if } u\in\{0,2,\dots,2k\} \\ \{(i,j)\in\{0,\dots,k\}^2 \mid u\le i+j\le u+1, i \text{ or } j \text{ are odd}\} & \text{if } u\in\{1,3,\dots,2k-1\}. \end{cases}$$

图5展示了如何将语言 $\{L_i^k \cap \hat{L}_j^k\}_{0\le i\le k, 0\le j\le k}$ 分组（通过进行语言析取），从而构成 L'^k 的 COCOA。$L_i^k \cap \hat{L}_j^k$ 中 i 和 j 均为偶数的语言构成了该 COCOA 的接受层，而它们之间的语言则被分组成 L'^k 的 COCOA 的拒绝层。

定理 5.7 的正确性证明用到了几个观察：

观察 5.8。设 w 是一个词，使得满足条件 $w\in L_i^k \cap \hat{L}_j^k$ 的唯一最大对为 $(i,\ j)$（其中 $(i,\ j)\in\Gamma_u^k$，对于某个 $0\le u\le 2k$）。我们有 $w\notin\mathcal{L}(C_{u+1}^k)$。

证明。如果 $(i,\ j)$ 是满足条件 $w\in L_i^k \cap \hat{L}_j^k$ 的唯一最大对 $(i,\ j)$，那么我们有 $w\notin L_{i+1}^k$，且 $w\notin \hat{L}_{j+1}^k$。

我们考虑两种情形，即 u 为偶数和 u为奇数。

在第一种情况下，我们有 $i+j=u$，i 和 j 均为偶数，并且我们正在寻找 Γ_{u+1}^k 中另一个满足 $i'\le i$（使得 w 被 $\hat{L}_{i'}^k$ 接受）且满足 $j'\le j$（使得 w 被 $L_{j'}^k$ 接受）的 (i',j')。如果 u 为偶数，那么 $u+1$ 为奇数，根据定理 5.7 中 $\{\Gamma_m^k\}_{0\le m\le 2k}$ 的定义，要使 (i',j') 包含于 Γ_{u+1}^k，必须有 $u+1\le i'+j'\le u+2$。由于 $i'\le i$、$j'\le j$ 和 $i'+j'\ge i+j+1$，我们得出 w 不被 C_{u+1}^k 接受，因为关于 i' 和 j' 的三项要求无法同时满足。

对于 u 为奇数的情况，我们有 $u \le i+j \le u+1$，并且 i 或 j 为奇数。这里，我们正在寻找满足 $u+1 = i'+j'$ 和 i' 的 $i' \le i$、$j' \le j$，且 j' 为偶数。这样一来，只有 $i = i'$ 和 $j = j'$ 可作为解；然而，这与 i 或 j 为奇数相矛盾，因为 i' 和 j' 为偶数。因此，我们得出 w 不被 C^k_{u+1} 接受。

□

观察 5.9。对于某些 $0 < i \le k$、$0 \le j \le k$ 和 $0 \le u \le 2k$，若对于偶数 i 和 j，有 $(i-1, j) \in \Gamma^k_{u-1}$，则 $(i, j) \in \Gamma^k_u$。

类似地，对于某些 $0 \le i \le k$、$0 \le j \le k$ 和 $0 \le u \le 2k$，若对于偶数 i 和 j，有 $(i, j-1) \in \Gamma^k_{u-1}$，则 $(i, j) \in \Gamma^k_u$。

证明。设 $(i-1, j) \in \Gamma^k_{u-1}$，其中 i 为偶数，且 j。那么，$i-1$ 为奇数，j 为偶数，而 $u-1$ 为奇数。因此 $u-1 \le i-1+j \le u$。这意味着 $u = i+j$，进而根据定理 5.7 中$\{\Gamma^k_m\}_{0 \le m \le 2k}$ 的定义，我们有 (i, j) 属于 Γ^k_u。

另一种情形可以用类似的方法证明。 □

观察 5.10。设 $(i, j) \in \Gamma^k_u$ 和 (i', j') 是满足 $0 \le i' \le i, 0 \le j' \le j, (i, j) \ne (i', j')$，且 i'和 j' 均为偶数的一对。于是有 $(i', j') \in \Gamma^k_0 \cup \ldots \cup \Gamma^k_{u-1}$。

证明。设 u' 是满足 $(i', j') \in \Gamma^k_{u'}$的元素。

若 u 为偶数，则 i 和 j 均为偶数，且 $u = i+j < i'+j' = u'$，因此 $(i', j') \in \Gamma^k_0 \cup \ldots \cup \Gamma^k_{u-1}$。

若 u为奇数，则 $u \le i+j \le u+1$。我们区分两种情形，即 $i+j = u$ 以及 $i+j = u+1$。

在第一种情况下，由于 $u' = i'+j'$（因为 u' 是偶数），$i' \le i$、$j' \le j$，以及 $(i, j) \ne (i', j')$，我们得到 $u' < u$，这已经足够。

在第二种情况下，我们有 $i+j = u+1$。由于 i 或 j 为奇数，且 u 为奇数，因此要使 $u+1$ 为偶数，i 和 j 都必须为奇数。由于 $i' \le i$、$j' \le j$，且 i'和 j' 都为偶数，因此我们有 $i' < i$ 和 $j' < j$。所以我们有 $u' = i'+j' \le i+j-2 = u-1$。于是，我们再次得到 $(i', j') \in \Gamma^k_0 \cup \ldots \cup \Gamma^k_{u-1}$。

□

定理 *5.7* 的证明。我们对 u 进行归纳法来证明该断言。具体而言，我们证明，自然颜色为 u 的词恰好是被 C^k_{u+1} 拒绝、但被 C^k_u 接受的那些词。注意，这等价于证明：对于每个自然颜色为 u 的词 w，对于满足 $0 \le i \le k$ 和 $0 \le j \le k$，且 $w \in L^k_i \cap \hat{}L^k_j$ 的唯一极大数对 (i, j)，我们有 $(i, j) \in \Gamma^k_u$。

我们将断言的证明分为四种情况，分别对应偶数 u/奇数 u以及证明的两个方向的所有可能组合。

偶数 u，$\Leftarrow$： 先考虑一个偶数 $u \ge 0$，并令 w 为一个被 C^k_u 接受、但被 C^k_{u+1}拒绝的词。唯一的最大对 (i, j) 满足 $w \in L^k_i \cap \hat{}L^k_j$，且$i+j = u$ 为偶数 i 和 j（因为 C^k_u 是由针对这类对的 $L^k_i \cap \hat{}L^k_j$的自动机析取构造而成的）。

由于 i 和 j 都是偶数，因此 w 既属于 L^k，也属于 $\hat{}L^k$，所以应该属于 $\mathcal{C}^k$的语言。由于 u 是偶数，且 w 被 C^k_u 接受、被C^k_{u+1}拒绝，确实如此。

现在我们来证明，w 满足具有自然颜色 u 的条件。具体而言，设 w' 是 w 的一个扩展，且(i', j')是满足$w' \in L^k_{i'} \cap \hat{}L^k_{j'}$的唯一最大对。根据引理 5.6，我们有 $i' \ge i$ 和 $j' \ge j$。若 $i = i'$ 且 $j = j'$，则所得的词仍属于 L^k、$\hat{}L^k$以及 $\mathcal{C}^k$的语言，因此满足以下条件：

w 具有自然颜色 u。然而，若 $i' < i$ 或 $j' < j$，则所得的词会被 C_u^k拒绝；根据归纳假设，它具有更小的自然颜色，这也满足词 w 具有自然颜色 u的条件。

偶数 u**，** $\Rightarrow$**:** 设 w 是一个自然颜色为 u 的词，且 u 为偶数。由于 u 为偶数，该词属于 L'^k，因此存在唯一的最大对 $(i,\ j)$，其 i 为偶数，且 j满足 $w \in L_i^k \cap L_j^k$。我们证明 $(i,\ j) \in \Gamma_u^k$。

我们首先区分 u 是否为 0。若 $u = 0$，则不存在某个层级 $(i',\ j')$，使得 i' 或 j' 为奇数，且满足 $i' \leq i$ 和 $j' \leq j$；因为根据引理 5.6，我们便能构造出一个不属于 L'^k 的词 w'，它是 w 的扩展，这与 w 的自然颜色为 0 矛盾。这意味着 $i = 0$ 和 $j = 0$，否则便可以（根据引理 5.6）扩展得到一个词 w'，使得 $(0,\ 1)$ 或 $(1,\ 0)$ 是 w' 的唯一最大对。因此，w 的唯一最大对为 $(0,\ 0)$，且此对包含于 Γ_0^k 中。

对于 $u \geq 2$ 的情形，我们有 $i > 0$ 或 $j > 0$。考虑 $i > 0$ 的情形。此时存在 w 的一个扩展 w'，其最大对为 $(i-1, j)$。由于 w 的自然颜色是 u，根据假设且因为 $w' \notin L^k$，w' 的自然颜色至多为 $u-1$。因此，根据归纳假设，我们有 $(i-1, j) \in \Gamma_0^k \cup \ldots \cup \Gamma_{u-1}^k$。如果 $(i-1, j) \in \Gamma_0^k \cup \ldots \Gamma_{u-3}^k$，那么根据观察 5.9 和归纳假设，我们会得到 w 的自然颜色为 $u-2$，这与假设 w 的自然颜色为 u 矛盾。我们也不可能有 $(i-1, j) \in \Gamma_{u-2}^k$，因为根据归纳假设，$w'$ 也会属于 L'^k，这与假设 i 和 j 都是偶数矛盾。因此，我们有 $(i-1, j) \in \Gamma_{u-1}^k$。再次应用观察 5.9，可得 $(i, j) \in \Gamma_u^k$。

$j > 0$ 的情形类似。

奇数 u**，** $\Leftarrow$**:** 现在考虑一个奇数 $u \geq 0$，并设 w 是一个被 C_u^k 接受、但被 C_{u+1}^k 拒绝的词。唯一的最大层级（$i,\ j$）满足 $w \in L_i^k \cap \hat{} L_j^k$，且 $u \leq i + j \leq u + 1$ 中的 i 或 j 为奇数（或两者皆为奇数）。

我们证明 w 的自然颜色是 u。为此，考虑一个扩展 w 的词 w'，并设（$i',\ j'$）是满足 $w' \in L_{i'}^k \cap L_{j'}^k$ 的唯一最大对。由引理 5.6 可知，$i' \leq i$ 且 $j' \leq j$。如果 i' 或 j' 仍为奇数，那么 $w' \notin L'^k$，于是词 w 具有自然颜色 u 的条件便得到满足。或者，根据观察 5.10，w' 被 C_u^k 拒绝；根据归纳假设，这足以证明 w' 的自然颜色至多为 $u-1$。因此，对于 w 的所有扩展 w'，所得词仍不属于L'^k，或者其自然颜色至多为 $u-1$，故断言得证。

奇数 u**，** $\Rightarrow$**:** 现在令 w 为一个自然颜色为 u 且 u 为奇数的词。因此，该词不属于 L'^k；对于满足 $w \in L_i^k \cap \hat{} L_j^k$ 的唯一最大值（$i,\ j$），我们有 i 为奇数或 j 为奇数，因为所取语言交集中的某个 COCOA 需要拒绝 w，该词才不会属于 L'^k。

我们需要证明（$i,\ j$）属于 Γ_u^k，因此该词属于 $\mathcal{L}(C_u^k)$。由于此时（$i,\ j$）属于 Γ_u^k，根据观察 5.8，我们有 w 不属于 $\mathcal{L}(C_{u+1}^k)$。

考虑通过剩余语言不变词注入从 w 得到的词集 w'。忽略 L'^k中的那些词 w'，因为无需证明它们的任何性质。根据假设，所有此类词的自然颜色都不大于 $u-1$。我们构造所有可能出现在任一此类词 $w' \in L'^k$中的指标对 Y 的集合。注意，所有此类指标对都属于 $\Gamma_0 \cup \Gamma_2 \cup \ldots \Gamma_{u-1}$（因为假设 w的自然颜色不大于 u）。令 $i_{max} = \max\{i' \in \mathbb{N} \mid \exists j'.(i',\ j') \in Y\}$，并令 $j_{max} = \max\{j' \in \mathbb{N} \mid \exists i'.(i',\ j') \in Y\}$。我们区分三种情形：

- i为奇数且 j为偶数：那么， $j = j_{max}$如$(i-1,\ j_{max}) \in Y$，且 $i = 1 + i_{max}$，因为任何值 $i > 1 + i_{max}$都会使 w可以扩展为包含在 $L^k_{i_{max}+2} \cap \hat{} L^k_j$中的词。这将与 u 是 w的自然颜色这一点矛盾，因为其唯一最大对为$(i_{max}+2,\ j)$的词可以扩展为$(i_{max}+1,\ j)$中的一个词，且该词在L'^k中的包含关系不同；又因为根据归纳假设，该词的颜色不能为 $u-1$，这意味着 w 的自然颜色不能为 u。

  由于 $(i_{max},\ j)$ 属于 Γ^k_{u-1}（关于 u，要具有自然颜色 u，所有在 L'^k 中包含关系不同的扩展所能具有的颜色至多为 $u-1$；而根据 $(i_{max},\ j)$，不可能具有更低的自然颜色，否则其自然颜色也会低于 $u-1$，这与 w 具有自然颜色 u 相矛盾），根据集合 $\{\Gamma^k_m\}_{0\le m\le 2k}$ 的定义，我们有 $(i_{max}+1,\ j) \in \Gamma^k_u$。
- i为偶数， j 为奇数：此情形与前一种情形类似。
- i 为奇数， j 为奇数。于是，$(i_{max}+1,\ j_{max}+1)$ 是唯一最大的元组，使得当 w' 的接受情况不同于 w 的接受情况时， w' 可以位于这一层（根据引理 5.6）。根据第一种情形中的相同推理，它不可能位于更大的元组处。根据假设， w 具有自然颜色 u，因此 w 所有在 L'^k 中包含关系不同的扩展都需要具有至多为 $u-1$ 的自然颜色，由此可知 $(i_{max},\ j_{max}) \in \Gamma^k_{u-1} \cup \ldots \cup \Gamma^k_1$。与此同时，$(i_{max}+1,\ j_{max}+1) \notin \Gamma^k_{u-1} \cup \ldots \cup \Gamma^k_1$，否则根据归纳假设， w 的自然颜色会低于 u。根据 $\{\Gamma^k_m\}_{0\le m\le 2k}$ 集合的定义，我们便有 $(i,\ j) = (i_{max}+1,\ j_{max}+1) \in \Gamma^k_u$。

□

定理 5.7 为如何根据针对 L^k 和 $\hat{} L^k$的 COCOA 构造 $\mathcal{C}^k$ 提供了蓝图。特别地，我们可以通过一系列析取和合取操作得到 $\mathcal{C}^k$ 。根据下文的命题 5.12，这一刻画使我们能够推断，$\mathcal{C}^k$ 的规模必定关于 k呈指数级。为简化构造和证明，我们将在确定性而非历史确定性的 co-B¨uchi 自动机上执行析取和合取操作，然后只证明所得自动机即使允许历史确定性，其规模仍然需要达到指数级。下述引理将确定性 B¨uchi 自动机进行合取或析取的常见结果推广到基于转移的接受条件（以及 co-B¨uchi 接受条件）的情形。

引理 5.11. 设 $\mathcal{A}_1, \ldots, \mathcal{A}_n$ 是定义在同一字母表上的确定性 *co-B ¨uchi* 自动机。我们可以构造一个确定性 *co-B ¨uchi* 自动机 $\mathcal{A}^\wedge$ ，识别这些语言合取的自动机，其规模为 $|\mathcal{A}_1| \cdot \ldots \cdot |\mathcal{A}_n|$；还可以构造一个确定性 *co-B ¨uchi* 自动机 $\mathcal{A}^\vee$，识别这些语言析取的自动机，其规模为 $|\mathcal{A}_1| \cdot \ldots \cdot |\mathcal{A}_n| \cdot n$。

证明。通过将现有的（民间）方法调整到基于转移的接受条件情形，可以证明这一断言。据作者所知，所得构造并未在现有文献的任何地方明确出现。不过，用于构造和操作自动机的 spot 框架 [DRC+22] 实现了自动机对的相关操作。

设对每个 $1 \le i \le n$，给定自动机 $\mathcal{A}_i$，其形式为元组 $(Q^i,\ \Sigma,\ \delta^i,\ q^i_0)$。我们定义 $\mathcal{A}^\wedge = (Q^\wedge,\ \Sigma,\ \delta^\wedge,\ q^\wedge_0)$，其中 $Q^\wedge = Q^1 \times \ldots \times Q^n$，$q^\wedge_0 = (q^1_0,\ \ldots,\ q^n_0)$；对于每个 $(q_1,\ \ldots,\ q_n)$、$(q'_1,\ \ldots,\ q'_n) \in Q^\wedge$、$x \in \Sigma$ 和 $c \in \{1,2\}$，若存在某些 $c_1,\ \ldots,\ c_n,\ c'$，使得对于每个 $1 \le i \le n$，都有 $(q_i,\ x,\ q'_i,\ c_i) \in \delta^i$，且 $c' = \max(c_1,\ \ldots,\ c_n)$，则有 $((q_1,\ \ldots,\ q_n)\ x,\ ((q'_1,\ \ldots,\ q'_n)\ c')) \in \delta^\wedge$。

注意，该自动机是确定性的，当且仅当每个输入的 co-B¨uchi 自动机都接受某个词时，它才接受该词。这是因为 $\mathcal{A}^\wedge$ 会并行模拟所有自动机

并且，当且仅当分量自动机中的所有转移都是接受转移时，它才有一条接受转移。因此，当且仅当对于某个词 w，在 w in $\mathcal{A}^\wedge$ 的某次运行中，最终只发生接受转移时，所有自动机都会接受。

我们还定义 $\mathcal{A}^\vee = (Q^\vee, \Sigma, \delta^\vee, q_0^\vee)$，其中 $Q^\vee = Q^1 \times \ldots \times Q^n \times \{1, \ldots, n\}$，$q^\vee = (q_0^1, \ldots, q_0^n, 1)$，并且对于每个 $(q_1, \ldots, q_n, j)$、$(q_1', \ldots, q_n', j') \in Q^\vee$，$x \in \Sigma$ 和 $c \in \{1, 2\}$，我们有 $((q_1, \ldots, q_n, j), x, ((q_1', \ldots, q_n', j'), c')) \in \delta^\wedge$，当且仅当存在某些 $c_1, \ldots, c_n, c'$，使得对于每个 $1 \leq i \leq n$，我们都有 $(q_i, x, q_i', x_i) \in \delta^i$；若 $j = n$ 且 $j' = 1$，则为 $c' = 1$，否则为 $c' = 2$；并且若 $c_j = 1$，则 $j' = (j \bmod n) + 1$，否则为 $j' = j$。

这个自动机也是确定性的。为了说明它接受 $\mathcal{A}_1, \ldots, \mathcal{A}_n$ 的语言并集，考虑词 w 被某个自动机 $\mathcal{A}_i$ 接受的情况。此时，沿着 w 对 $\mathcal{A}^\vee$ 的运行，计数器（即状态集的最后一个元素）最终会停留在值 i，之后不再发生拒绝转移。另一方面，如果所有自动机 $\mathcal{A}_1, \ldots, \mathcal{A}_n$ 都拒绝词 w，那么沿着 w 的运行，计数器会无限多次遍历所有可能的值；并且，由于计数器从 n 切换到 1 时会发生拒绝转移，自动机 $\mathcal{A}^\vee$ 会拒绝该词。

□

我们在下面命题的证明中使用这个引理：

命题 5.12。 设 $\mathcal{C}^k = (C_1^k, \ldots, C_{2k}^k)$ 是 $L^k \cap \hat{} L^k$ 的一个 *COCOA*。

对于某个 $1 \leq i \leq 2k$，语言 $\mathcal{L}(C_i^k)$ 的最小确定性 *co-B ̈uchi* 自动机至多有 $2^{2k} \cdot k$ 个状态。对于偶数 k，*HD-tCBW* C_k^k 至少有 2^k 个状态。对于奇数 k，C_{k-1}^k 至少有 2^{k-2} 个状态。

证明。 对于所有语言 L_i^k 和 $\hat{} L_j^k$，分别将对应的两个状态确定性自动机记为 A_i^k 和 $\hat{} A_j^k$。根据引理 5.2，它们各有两个状态。

对于第一部分，注意到对于所有 $0 \leq i \leq 2k$，集合 Γ_u^k 至多有 k 个非支配元素，即集合中不存在另一个不同的数对 (i', j')，使得 $i' \geq i$ 且 $j' \geq j$ 的数对 (i, j)。构造 $C_u^k = \bigcup_{(i,j) \in \Gamma_u^k} A_i^k \cap \hat{} A_j^k$时，只需考虑非支配数对，因为对于一个受支配数对 (i', j')，所有被 $A_{i'}^k \cap \hat{} A_{j'}^k$ 接受的单词，也都会被某个非支配数对 (i, j)对应的 $A_i^k \cap \hat{} A_j^k$ 接受。根据引理 5.11，对于某个数对 (i, j)，确定性 co-B ̈uchi 自动机 $A_i^k \cap \hat{} A_j^k$ 只需 4 个状态。根据同一引理，将 k 个这样的自动机取并，得到的自动机至多有 $2^{2k} \cdot k$ 个状态。

对于第二部分，考虑 $C_k^k = \bigcup_{(i,j) \in \Gamma_k^k} A_i^k \cap \hat{} A_j^k$ 在 k 为偶数时的情况。这里，我们对 $\frac{k}{2}$ 个 4 状态自动机取析取,_ 得到一个至多有 2^k 个状态的自动机。同时，这也是下界，因为 $\mathcal{L}(C_k^k)$ 的剩余语言数量为 2^k。这是因为，对于每个满足 $(i, j) \in \Gamma_u^k$ 的语言 $L_i^k \cap \hat{} L_j^k$，都有一个词后缀 $\tilde{w} = (a_{4k-2i+1} a_{2j-2})^\omega$，它不会被任何满足 $(i', j') \in \Gamma_u^k$ 的 $L_{i'}^k \cap \hat{} L_{j'}^k$ 所接受；同样，对于 $(i, j) \neq (i', j')$，该词后缀仅属于那些前缀中字母 X_i 和 Y_j 各出现偶数次的前缀所对应的 $L_i^k \cap \hat{} L_j^k$。这意味着，对于 $(i, j) \in \Gamma_k^k$，一个 HD-tCBW 对于字母 X_i 或 Y_j 出现次数不同的词，具有不同的剩余语言。因此，剩余语言的数量为 2^k。由于最小规范 HD-tCBW 在语义上是确定的，[ARK22], 若具有 2^k 个剩余语言，则意味着其大小下界为 2^k，适用于 C_k^k。k 为奇数时的情况类似。

□

最后我们讨论一下，将 L^k 和 $\hat{}L^k$ 表示为确定性奇偶自动机时，不会出现类似的膨胀。

命题 5.13. 语言 L^k （来自定义 *5.1*）和 $\hat{}L^k$ （来自定义 *5.5*）都可以表示为具有 2^k 个状态的确定性奇偶自动机（且状态数不能更少）。

存在一个用于 $L^k \cap \hat{}L^k$ 的确定性奇偶自动机，其状态数不超过 $2^{4k^2} \cdot k^{2k}$。

证明。 我们可以构造一个用于 L^k 的确定性奇偶字自动机（DPW） $\mathcal{P}^k = (Q, \Sigma, \delta, q_0)$，其中 $Q = \mathbb{B}^k$, $q_0 = (0, \ldots, 0)$，并且对于所有 $(b_1, \ldots, b_k) \in Q$ 和 $x \in \Sigma$，若 $x = X_i$，则有 $\delta((b_1, \ldots, b_k)\ x) = ((b'_1, \ldots, b'_k)\ c)$；否则，对于所有 $1 \leq i \leq k$，有 $b'_i = b_i$。该转移中 c 的值定义如下：

$$c = \begin{cases} 0 & \text{if } x \in \{X_1, \ldots, X_k, Y_1, \ldots, Y_k\} \\ i & \text{if } x = a_{4k-2i} \text{ for some } 1 \leq i \leq k \\ i & \text{if } x = a_{4k-2i+1} \text{ and } b_i = 0 \text{ for some } 1 \leq i \leq k \\ i-1 & \text{if } x = a_{4k-2i+1} \text{ and } b_i = 1 \text{ for some } 1 \leq i \leq k \\ k & \text{otherwise.} \end{cases}$$

我们证明，该自动机以其自然颜色识别每个词（相对于 L^k）。为此，对于某个词 w，我们将 $\mathcal{P}^k$ 的一次运行所诱导的颜色与该词所属的 L^k_1、.....、L^k_k 语言进行比较。

• 如果 $\{X_1, \ldots, X_k, Y_1, \ldots, Y_k\}$ 中有无穷多个字母出现，那么该词不属于 L^k_1。 $\mathcal{P}^k$ 以颜色 0 识别该词，而这正是该词的自然颜色（相对于 L^k）。

• 如果对于某个 $1 \leq i \leq k$，字母 a_{4k-2i} 出现无穷多次，但对于某个指标 $j < 4k-2i$，字母a_j只出现有限次，那么该词属于 L^k_i，但不属于 L^k_{i+1}。在这种情况下， $\mathcal{P}^k$ 以颜色 i 识别该词。

• 如果对于某个 $1 \leq i \leq k$，字母 $a_{4k-2i+1}$ 出现无穷多次，但对于某个指标 $j < 4k-2i+1$，字母 a_j 只出现有限次，并且 X_i 个字母的数量为偶数，那么该词属于 L^k_i，但不属于 L^k_{i+1}。在这种情况下， $\mathcal{P}^k$ 以颜色 i 识别该词。

• 如果对于某个 $1 \leq i \leq k$，字母 $a_{4k-2i+1}$ 出现无穷多次，而对于某个索引 $j < 4k-2i+1$，字母a_j只出现有限次，并且 X_i 字母的数量为奇数，那么该词属于 L^k_{i-1}，但不属于 L^k_i。 $\mathcal{P}^k$以颜色 $i-1$ 识别该词。

• 否则，该词最终只由满足 $j < 4k-2k$的字母a_j 组成。此时，即使 L^k_k也接受该词，而且该词的自然颜色为 k。我们有 $\mathcal{P}^k$ 以颜色 k识别该词。

请注意， $\mathcal{P}^k$ 是识别 L^k 的最小确定性奇偶自动机，因为它有 2^k 个状态，而 L^k 的剩余语言数量为 2^k，所以它不可能更小。

可以从$\hat{}L^k$ 构造出一个类似的确定性奇偶字自动机，只需将 X_i 字符替换为 Y_i，并重新编号a_j 字母的索引。

对于 $L^k \cap \hat{}L^k$ 的确定性奇偶字自动机，我们采用命题 5.12，为 L^k 的 COCOA 的每个层级构造大小至多为 $2^{2k} \cdot k$ 的确定性 co-B¨uchi 自动机；然后按照命题 4.2，将这些确定性自动机构造为乘积奇偶自动机，从而得到一个用于 L'^k、大小至多为 $(2^{2k} \cdot k)^{2k} = 2^{4k^2} \cdot k^{2k}$ 的确定性奇偶自动机。 □

命题 5.13、命题 5.12 和引理 5.2 共同表明，虽然 L^k 和 $\hat{}L^k$可以用 COCOA 表示，且其简洁程度呈指数级，优于表示这些语言的任何确定性

奇偶自动机，但在计算 $L^k \cap \hat{} L^k$ 的 COCOA 时，指数简洁性便不复存在。

注 5.14. 取两个 COCOA 的析取（而非合取）也会失去指数简洁性。

证明。 通过在链中添加一个接受通用语言的 HD-tCBW 作为新的首个自动机，并将所有链元素后移一位，可以得到分别识别 L^k 和 $\hat{} L^k$ 所对应语言的补集的 COCOA。对所得的识别补语言的 COCOA 取合取后，根据定理 5.7 的构造，可得到一个首个自动机接受通用语言的 COCOA。将其移除后，即得到识别 $L^k \cup \hat{} L^k$ 的 COCOA。对合取自动机应用命题 5.12，可得识别 $L^k \cup \hat{} L^k$ 的 COCOA 的指数级规模下界。对于用于比较的奇偶自动机，补运算可以通过将每个转移颜色加 1 来实现，且不会造成膨胀。

□

6. Complementing COCOA

在本节中，作为本文的最后一项贡献，我们研究 co-B¨uchi 自动机链的补运算问题。我们证明，这种补运算需要重新构造，并改变识别这些词所用的自然颜色。可能出现这样的情况：两个词 w 和 w' 都由某个颜色为 i的 COCOA 识别，但用于补集的 COCOA 会以颜色 $i-1$ 识别 w，同时以颜色 $i+1$识别 w'。因此，与确定性奇偶自动机不同，对 COCOA 求补并非简单地将每个转移颜色加一即可完成，而是一项非平凡的操作。更进一步，正如我们在本文本节中所证明的，补运算可能导致 COCOA 的规模呈指数级增长。

我们定义一个 COCOA 族 $\mathcal{C}^1$，$\mathcal{C}^2$，……，其中每个 COCOA $\mathcal{C}^k$ 总共有 $(4+3\cdot k)$ 个状态，而用于识别 $\mathcal{C}^k$ 的补集的 COCOA 中，第一个自动机至少需要2^k 个状态。图 6 展示了 COCOA $\mathcal{C}^k$ 的 co-B¨uchi 自动机。

语言 L^k 的字母表 Σ 中有三组字母，分别由 COCOA $\mathcal{C}^k$ 表示，即 $\{a_1, \ldots, a_{2k+1}\}$、$\{X_1, \ldots, X_k\}$ 和 $\{Y_1, \ldots, Y_k\}$。对于 $i \geq 2$，每个自动机 A_i^k 都接受满足以下条件的单词：来自 X_1、……、X_k 的字母仅出现有限次，来自 $\{a_{2k-i+5}, \ldots, a_{2k+1}\}$ 的字母仅出现有限次，并且 a_{2k-i+4} 仅出现有限次，或者在第一个 Y_{i-1} 字母之后不再出现字母 X_{i-1}。因此，如果 Y_{i-1}和 X_{i-1} 按该相对顺序出现在一个单词中，那么沿着一个被接受的单词可能无限次出现的、来自 a_1、……、a_{2k+1} 的字母集合就会发生变化。自动机 A_1 接受所有来自 X_1、……、X_k 的字母仅出现有限次的单词。

与上一节一样，我们需要证明对于每个 $k \in \mathbb{N}$, $\mathcal{C}^k$ 都是一个有效的 COCOA，即它会根据自动机链所表示的语言，为每个词识别其自然颜色。

引理 6.1. 设 $\mathcal{C}^k = (A_1^k, \ldots, A_{k+1}^k)$ 是图 *6* 中给出的自动机序列（对某个 $k \in \mathbb{N}$ 而言），且 $L^k = \mathcal{L}(A_1^k, \ldots, A_{k+1}^k)$。

我们有 $\mathcal{C}^k$ 是一个有效的 *COCOA*，也就是说，对于 $1 \leq i \leq k+1$，每个自动机 A_i^k 都恰好接受相对于 L^k 具有至少 i 个自然颜色的词。

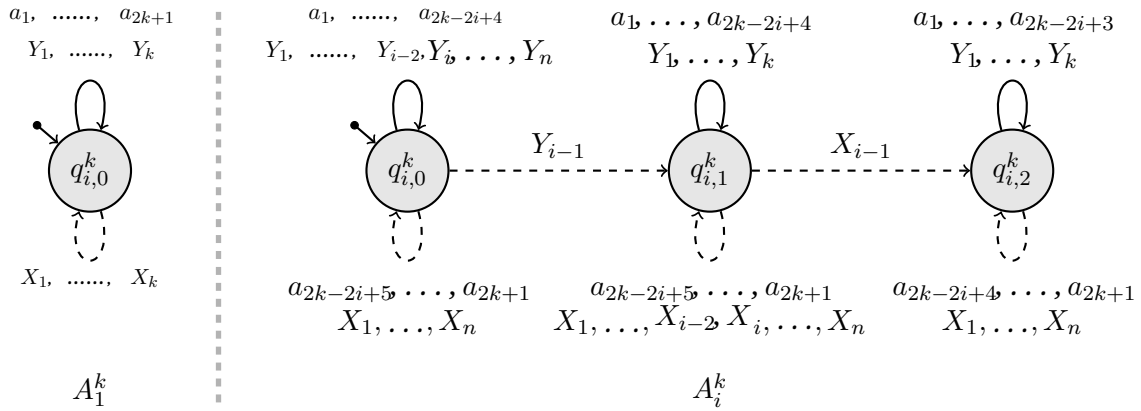

图 6：用于第6节中展示补运算可能导致指数级膨胀的 COCOA $\mathcal{C}^k = (A^k_1, \ldots\ldots, A^k_{k+1})$ 的 Co-B¨uchi 自动机 A^k_1 和 A^k_i（针对 $2 \leq i \leq k+1$）（针对某个 $k \in \mathbb{N}$）

证明。 首先注意，链中的所有自动机都是历史确定性的，因为它们也是确定性的。还要注意，对于每个 $1 \leq i \leq k$，都有 $\mathcal{L}(A^k_i) \supset \mathcal{L}(A^k_{i+1})$，因为链中的所有自动机只接受最终以 $\{Y_1, \ldots, Y_k, a_1, \ldots, a_{2k+1}\}$ 中的字母结尾的词，而可能无限次出现的字母集合 a_1、……、a_{2k+1} 会随着层级递增而严格缩小。

现在我们通过归纳法证明，对于每个 $1 \leq i \leq k$，语言 L^k_i 恰好包含那些关于 L^k 的自然颜色至少为 i的词。

归纳基础： 设 w 是一个词。我们需要证明，恰好所有满足以下条件的词 $w \in L^k$ 才是那些不属于语言 A^k_1的词：每个导向某个词 w'的剩余语言不变的词注入序列，都使得该词也具有 $w' \in L^k$ 。

$\Rightarrow$：设 w 是一个不属于语言 A^k_1的词。这意味着，$X_1, \ldots\ldots, X_k$ 中的某个字母在该词中无穷多次出现。注入额外的字母不会改变这一点，因此，对于所得的词，我们有它属于 L^k ，并且也不属于语言 A^k_1 。

$\Leftarrow$：设 $w \in L^k$ 是一个词，对于某个 J，所有剩余语言不变的词注入序列都会导出一个同样属于 L^k的词。那么，这也包括通过无穷多次注入 a_{2k+1} 得到的词；这不会改变该词是否被 A^k_1 接受，却能确保它被 A^k_2、……、A^k_{k+1}拒绝。此外，相对于 L^k 而言，这种注入保持剩余语言不变，因为 $\mathcal{C}^k$ 中的所有 co-B¨uchi 自动机在读到这个字母时总会自环。因此，如果注入前 w 属于 L^k ，那么它早已被 A^k_1拒绝。

归纳步骤： $\Rightarrow$：设 w 是一个不属于 $\mathcal{L}(A^k_i)$ 的词。我们需要证明 w 具有至多为 $i-1$的自然颜色。如果该词不属于 $\mathcal{L}(A^k_{i-1})$，则由归纳假设可得结论。因此，只需考虑 A^k_{i-1} 接受该词的情形。

那么，我们可以得出：该词要么包含无限多个 $a_{2k-2i+5}$ 字母，要么在第一个 Y_{i-1} 字母之后包含 X_{i-1} 字母，且包含无限多个。注入任何残余语言不变词都无法改变这一点。在此语境下，请注意，如果该词包含一个 Y_{i-1} 字母，且其后没有 X_{i-1} 字母，那么在第一个 Y_{i-1} 字母之后注入X_{i-1} 字母就不是后缀语言不变的，因为这会改变 $(a_{2k-2i+4})^\omega$ 是否属于剩余语言。如果现在注入一个包含来自$a_{2k-2i+6}$、……、 a_{2k+1} 的字母且这些字母出现无限多次的词，那么该词就会被 A^k_{i-1}拒绝，由此根据

归纳假设，所得词的自然颜色至多为 $i-2$。否则，该词仍会被 A^k_{i-1} 接受并被 A^k_i拒绝，同时保持该词是否属于 L不变。对于所有残余语言不变注入，我们现在知道，所得词要么属于 L 当且仅当注入前它属于 L，要么所得词的自然颜色至多为 $i-2$；至此，归纳步骤的这一方向就完成了。

$\Leftarrow$: 设 w 是一个自然颜色至多为 $i-1$ 的词。我们需要证明 w 不属于 $\mathcal{L}(A^k_i)$。如果该词的颜色至多为 $i-2$，那么根据归纳假设，我们知道它不属于 $\mathcal{L}(A^k_{i-1})$，因此由链元素之间的严格语言包含关系，断言已经成立。因此，假设 w 的自然颜色恰好为 $i-1$。任何会改变该词是否属于 L^k 的后缀语言不变字注入，都必须使其导向一个颜色至多为 $i-2$ 的词。

用反证法，假设 w 被 A^k_i 接受。现在将 $a_{2k-2i+3}$ 无穷多次注入 w。所得词仍属于 $\mathcal{L}(A^k_i)$，但会被 $\mathcal{L}(A^k_{i+1})$ 拒绝，因此它属于 L 当且仅当 i 是偶数。由于它的自然颜色并非至多为 $i-2$，而且结果词的自然颜色也不可能是 $i-1$，因为 $i-1$ 的奇偶性与该词是否属于 L 不相符，这就证明结果词的自然颜色至少为 i；而如果 w 的自然颜色恰好为 $i-1$，这是不可能发生的。 □

现在来考虑对 COCOA $\mathcal{C}^k$ 取补集的问题。下述定理表明，表示 $\mathcal{C}^k$ 的补集的 COCOA，其第一个元素所需的状态数相对于 k 呈指数级增长。

定理 6.2. 设对于某个 $k \in \mathbb{N}$，$\mathcal{C}^k = (A^k_1, \ldots, A^k_{k+1})$ 是一条 *co-B ̈uchi* 自动机链，其结构如图 *6* 所示，并且 $L^k = \mathcal{L}(A^k_1, \ldots, A^k_{k+1})$。

编码 $\Sigma^\omega \setminus L^k$ 的 *COCOA* 的每个首个链元素至少需要 2^k 个状态。

证明。 我们通过刻画自然颜色恰为 0 的词来证明。我们将这些词构成的语言记为 $\hat{}L^k_0$，并证明它至少有 2^k 个剩余语言。由于用于 $\Sigma^\omega \setminus L^k$ 的 COCOA 的第一个元素恰好接受 $\hat{}L^k_0$ 的补集，并且经最小化的 co-B ̈uchi 自动机不失一般性地是语言确定的（因为 Abu Radi 和 Kupferman [ARK22] 的最小化算法会生成此类自动机），这就证明了用于 $\Sigma^\omega \setminus L^k$ 的 COCOA 的第一个元素至少需要 2^k 个状态。我们可以如下刻画 $\hat{}L^k_0$：

$$\hat{L}^k_0 = \{w_0 w_1 \ldots \in \Sigma^\omega \mid \forall 1 \le i \le k. \exists j \in \mathbb{N}. Y_i \notin \{w_0, \ldots, w_{j-1}\} \wedge w_j = Y_i \wedge X_i \notin \{w_{j+1}, w_{j+2}, \ldots\}, \exists_\infty j \in \mathbb{N}. w_j \in \{a_{2k}, a_{2k+1}\}\}$$

在这个等式中，表达式 $\exists_\infty j \in \mathbb{N}$ 表示存在无穷多个具有所述性质的元素$j \in \mathbb{N}$。为证明 L^k_0 恰好包含 $\Sigma^\omega \setminus L^k$中自然颜色为 0 的单词，考虑 $\hat{}L^k_0$ 中的一个词。它包含 $Y_1, \ldots, Y_k$ 中的所有字母，但在对应的 Y_i 字母之后不包含 $X_0, \ldots, X_n$ 中的任何字母。因此， A^k_1 接受该词，而 A^k_2拒绝该词，故该词属于 $\Sigma^\omega \setminus L^k$。

现在考虑一列注入点 J，它们都位于$\{Y_1, \ldots, Y_n\}$中的每个字母至少出现一次之后。对于 $2 \le i \le k+1$，任何自动机 A^k_i的任何运行，在所有 Y_i 个字母都出现后，都会处于状态$q^k_{i,\,1}$。任何后续字母 X_i （对于$1 \le i \le k$）都会改变 COCOA 的剩余语言（因为它会改变剩余词（$a_{2k+2i+4}$）$^\omega$ 是否被接受），因此，位于 J 中的剩余语言不变的词注入不能包含 X_i 个字母。如果不注入任何 X_i 字母，所得

词仍会被 A_1^k 接受，并仍会被 A_2^k拒绝，因此该词又包含于$\Sigma^\omega \setminus L^k$ 中。

对于反向，考虑一个词 w in $\Sigma^\omega \setminus L^k$，使得存在一个序列 J，其中所有剩余语言不变的词注入也都属于 $\Sigma^\omega \setminus L^k$。我们区分三种情形：

• 某个字母 Y_{i-1} 可能不属于 w。那么，沿着A_i^k 在 w上的运行，$q_{i,\,1}$ 不会被到达；由于注入 X_{i-1} 会使 COCOA 中自动机的所有自动机运行都进入自环，因此我们知道该注入是剩余语言不变的。然而，无穷多次注入 X_{i-1}会导致所得词相对于 L^k具有自然颜色 0，因此它不属于 $\Sigma^\omega \setminus L^k$，这就产生了矛盾。

• 或者，某个字母 X_{i-1} 在 w中第一次出现 Y_{i-1} 之后出现。但这样一来，A_i^k 在 w 上的（唯一）运行会在字母 X_{i-1}出现后转移到状态 $q_{i,\,2}^k$，使得此后对 X_{i-1} 的注入都是剩余语言不变的。按照上一种情形中的相同推理，我们得到矛盾。

• 最后，对于所有 $1 \le i \le k$，字母 X_i 不会出现在 w中最后一个 Y_i字母之后，并且对于所有 $1 \le i \le k$，这些 Y_i 字母都能在该词的某处找到。在这种情况下，每个生成的词在 $\mathcal{C}^k$ 中都具有至少为 1 的自然颜色（因为 A_1^k 接受它）。如果不存在这样的注入能改变该词相对于 L^k具有偶数自然颜色这一事实，那么 A_2^k 接受它而 A_3^k 拒绝它；否则，通过无限多次注入 a_{2k-2}，我们就能使生成的词不属于 L^k，或者该词实际上属于 L^k（因此不属于 $\Sigma^\omega \setminus L^k$）。因此，$a_{2k}$ 或 a_{2k+1} 必定已在 w中无限多次出现（因为该词被 A_3^k拒绝），从而使它属于 $\hat{}L_0^k$。

现在我们知道，$\hat{}L_0^k$ 至少有 2^k 个剩余语言：所有仅使用 $\{Y_1, \ldots, Y_k, a_1, \ldots, a_{2k+1}\}$ 中的字母的词，只要它们在迄今出现过的 $\{Y_1, \ldots, Y_k\}$ 中字母方面有所不同，其剩余语言就不同；具体而言，它们的剩余语言中属于 $(\{Y_1, \ldots, Y_k\})^*\{a_1, \ldots, a_{2k+1}\}^\omega$ 的元素也有所不同。

由于针对 $\Sigma^\omega \setminus \hat{}L_0^k$ 的最小 co-B¨uchi 自动机（它是针对 $\Sigma^\omega \setminus L^k$ 的最小化 COCOA 中的第一个 co-B¨uchi 自动机）可以不失一般性地假定会跟踪剩余语言，因此针对 $\Sigma^\omega \setminus L^k$ 的第一个 COCOA 自动机至少需要有 2^k 个状态。 □

我们注意到，与合取和析取运算类似，证明 COCOA 补运算会导致指数级膨胀的定理，利用了 COCOA 的这一性质：其剩余语言的数量可能达到状态总数的指数级。

7. 结论

本文通过汇总关于具有基于转移的接受条件的历史确定性 co-B¨uchi 自动机链（COCOA）的已有结果，从中推导出推论，并以三项并非现有工作的推论的新技术结果补充了由此形成的 COCOA 已知简洁性性质综述，从而深入考察了这类链的简洁性。

特别是，我们证明，将待表示的语言划分为多个层级后，即使各个层级上的自动机并不简洁，COCOA 仍可能比确定性奇偶自动机简洁得多，甚至达到指数级差距。其次，我们证明，通过执行语言析取或合取操作，指数简洁性可能会彻底丧失。最后，我们证明，即使对单个 COCOA 取补，也可能需要指数级

膨胀。第一个结果甚至适用于只有一种剩余语言的语言；而第二个和第三个结果中简洁性的损失，则是因为所得 COCOA 中的某个 co-B¨uchi 自动机需要表示指数数量的剩余语言。后两项技术结果均不依赖于历史确定性 co-B¨uchi 自动机相对于确定性 co-B¨uchi 自动机的简洁性。

除了让我们得以深入了解 COCOA 作为 ω-正则语言表示形式的能力与局限之外，我们的研究结果也为今后开发 COCOA 运算算法，以及将 COCOA 用于实际应用的工作提供了参考。具体而言，合取／析取／补运算的下界为今后计算 COCOA 布尔组合的算法提供了比较基准。此外，由于所有关于 COCOA 布尔组合的研究结果中，规模呈指数级增长的原因都是 COCOA 可能具有指数数量的剩余语言，我们的结果促使人们探索用于 ω-正则语言的替代表示形式。这类表示形式既能继承 COCOA 的优良特性（如多项式时间最小化），又无法表示剩余语言数量相对于状态数呈指数级增长的语言。这样的模型或许能够支持布尔运算，同时只需多项式级的表示膨胀，并且还可能让执行此类运算的算法更加简单。

References

[ARE25] Bader Abu Radi 和 R¨udiger Ehlers。刻画可在多项式时间内最小化的 ω-自动机。*CoRR*，abs/2504.20553，2025年。doi:10.48550/ARXIV.2504.20553。[ARK22] Bader Abu Radi 和 Orna Kupferman。GFG 转移型自动机的最小化与规范化。 *Log. Methods Comput. Sci.*，18(3)，2022年。doi:10.46298/LMCS-18(3:16)2022。[BCJ18] Roderick Bloem、Krishnendu Chatterjee 和 Barbara Jobstmann。图博弈与反应式综合。载于 Edmund M. Clarke、Thomas A. Henzinger、Helmut Veith 和 Roderick Bloem 编，*Handbook of Model Checking* 921 962 2018 10 1007 978 3 319 10575 8_，页码 – 。27 Springer，。doi: 。/ - - - - 。[BKR10] Udi Boker、Orna Kupferman 和 Adin Rosenberg。Büchi 自动机中的交替消除。载于 Samson Abramsky、Cyril Gavoille、Claude Kirchner、Friedhelm Meyer auf der Heide 和 Paul G. Spirakis 编，*37th International Colloquium on Automata Languages and Programming* ，，*(ICALP)* 6199 *Lecture Notes in Computer Science* 76 87 2010 第 卷，页码 – 。Springer，。doi:10.1007/978-3-642-14162-1_7。[Bok18] Udi Boker。为什么选择这些自动机类型? 载于 *22nd International Conference on Logic for Program-ming Artificial Intelligence and Reasoning (LPAR-22.)* 143 163, 2018 10 29007 3 ，，页码 – 。doi: 。/C BJ。[DRC+22] Alexandre Duret-Lutz、Etienne Renault、Maximilien Colange、Florian Renkin、Alexandre Gbaguidi Aisse、Philipp Schlehuber-Caissier、Thomas Medioni、Antoine Martin、J´erˆome Dubois、Cl´ement Gillard 和 Henrich Lauko。从 spot 2 0 2 10 *34th International* . 到 spot . : 有哪些新内容? 载于 ，，页码 – 。doi: 。/ 978-3-031-13188-2_9。[Ehl25a] R¨udiger Ehlers。协同 Büchi 自动机链有多简洁? 载于 *Sixteenth International Sympo-sium on Games*， *Automata*， *Logics*，*and Formal Verification (GandALF 2025)*，页码111–126，2025年9月。doi:10.4204/EPTCS.428.9。[Ehl25b] R¨udiger Ehlers。重新铺轨自动机。 *CoRR*，abs/2503.08438，2025年。doi:10.48550/ARXIV.2503. 08438。[EK24a] R¨udiger Ehlers 和 Ayrat Khalimov。完全广义的 reactivity(1) 综合。载于 Bernd Finkbeiner 和 Laura Kov´acs 编， *30th International Conference on Tools and Algorithms for the Construction and Analysis of Systems (TACAS)* 14570 *Lecture Notes in Computer* ，第 卷，页码 – 。*Science* 83 102 2024 10 1007 978 3 031 57246 3_6 Springer，。doi: 。/ - - - - 。[EK24b] R¨udiger Ehlers 和 Ayrat Khalimov。从线性时序逻辑到奇偶与 COCOA 的自然着色转换。 *CoRR*，abs/2410.01021，2024年。doi:10.48550/ARXIV.2410.01021。

[EKRS17] Javier Esparza、Jan Kret´ınsk´y、Jean-Fran˛cois Raskin 和 Salomon Sickert。从线性时序逻辑和极限确定性 Büchi 自动机到确定性奇偶自动机。载于 Axel Legay 和 Tiziana Margaria 编，*Tools and Algorithms for the Construction and Analysis of Systems - 23rdInternational Conference (TACAS)*，第10205卷 *Lecture Notes in Computer Science*，页码635–659，2017年。doi:10.1007/978-3-662-54577-5_25。[ES22a] R¨udiger Ehlers 和 Sven Schewe。无限词的自然颜色。载于 *42nd IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS)*，第250卷 *LIPIcs*，页码36:1–36:17。Schloss Dagstuhl - Leibniz-Zentrum für Informatik，2022年。doi:10.4230/LIPICS.FSTTCS.2022.36。[ES22b] R¨udiger Ehlers 和 Sven Schewe。无限词的自然颜色——完整演示，2022年12月。doi:10.5281/zenodo.15585670。[FZ17] Peter Faymonville 和 Martin Zimmermann。参数化线性动态逻辑。 *Inf. Comput.*，253 237 256，2017 10 1016 2016 07 009 : – 。doi: 。/J.IC. . . . [GV13] Giuseppe De Giacomo 和 Moshe Y. Vardi。有限轨迹上的线性时序逻辑与线性动态逻辑。载于 Francesca Rossi 编，*23rd International Joint Conference on Artificial Intelligence (IJCAI)* 854 860 2013 ，页码 – 。

IJCAI/AAAI，。URL: http://www.aaai.org/ocs/index. php/IJCAI/IJCAI13/paper/view/6997。[KMM06] Orna Kupferman、Gila Morgenstern 和 Aniello Murano。ω-正则自动机的类型性。*Int. J. Found. Comput. Sci.*，17(4):869–884，2006年。doi:10.1142/S0129054106004157。[KR10] Orna Kupferman 和 Adin Rosenberg。将线性时序逻辑转换为确定性自动机时的状态数膨胀。载于 Ron van der Meyden 和 Jan-Georg Smaus 编，*6th International Workshop on ModelChecking and Artificial Intelligence (MoChArt)*，第6572卷 *Lecture Notes in Computer Science*，页码85–94。Springer，2010年。doi:10.1007/978-3-642-20674-0_6。[KS15] Denis Kuperberg 和 Michal Skrzypczak。关于游戏友好自动机的确定化。载于 Magn´us M. Halld´orsson、Kazuo Iwama、Naoki Kobayashi 和 Bettina Speckmann 编，*42nd International Colloquium on Automata, Languages, and Programming (ICALP)*，第9135卷 *Lecture Notes in Computer Science*，页码299–310。Springer，2015年。doi: 10.1007/978-3-662-47666-6_24。[MH84] Satoru Miyano 和 Takeshi Hayashi。*ω Theoretical Computer*-词上的交替有限自动机。，( ): – 。doi: 。/ - ( ) - 。*Science* 32 3 321 330, 1984 10 1016 0304 3975 84 90049 5 。[Pnu77] Amir Pnueli。程序的时序逻辑。载于 *18th Annual Symposium on Foundations of Computer Science* 46 57 1977 10 1109 1977 32 ，页码 – 。IEEE 计算机学会，。doi: 。/SFCS. . . [Sch10] Sven Schewe。超最小化之外——最小化确定性Büchi自动机和确定性奇偶自动机是NP完全的。载于 *IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS)* 8 *LIPIcs* 400 411 ¨ ，第 卷，页码 – 。Schloss Dagstuhl - Leibniz-Zentrum für Informatik，2010年。doi:10.4230/LIPICS.FSTTCS.2010.400。

本作品采用知识共享署名许可协议授权。要查看该许可协议的副本，请访问 https://creativecommons.org/licenses/by/4.0/，或致函知识共享，第二街171号，300室，旧金山，加利福尼亚州 94105，美国；也可寄信至德国柏林10777，艾森纳赫街2号。